(12) INTERNATIONAL APPLICATION PUBLISHED UNDER THE PATENT COOPERATION TREATY (PCT)

(19) World Intellectual Property Organization
International Bureau

(43) International Publication Date
18 April 2002 (18.04.2002)

PCT

(10) International Publication Number
WO 02/30981 A1

(51) International Patent Classification[7]: C07K 14/705, 16/28, C12N 15/12, A01K 67/02, C12P 21/00, C12Q 1/68, A61K 38/17

(21) International Application Number: PCT/US01/07832

(22) International Filing Date: 13 March 2001 (13.03.2001)

(25) Filing Language: English

(26) Publication Language: English

(30) Priority Data:
09/684,393 10 October 2000 (10.10.2000) US

(71) Applicant: PE CORPORATION (NY) [US/US]; Millman, Robert, A., 761 Main Avenue, Norwalk, CT 06859 (US).

(72) Inventors: WEI, Ming-Hui; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). ZHONG, Wenyan; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). KETCHUM, Karen, A.; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). DIFRANCESCO, Valentina; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US). BEASLEY, Ellen, M.; c/o Celera, 45 West Gude Drive, Rockville, MD 20850 (US).

(74) Agent: MILLMAN, Robert, A.; Celera Genomics Corp., 45 West Gude Drive C2-4, Rockville, MD 20850 (US).

(81) Designated States *(national)*: AE, AG, AL, AM, AT, AU, AZ, BA, BB, BG, BR, BY, BZ, CA, CH, CN, CR, CU, CZ, DE, DK, DM, DZ, EE, ES, FI, GB, GD, GE, GH, GM, HR, HU, ID, IL, IN, IS, JP, KE, KG, KP, KR, KZ, LC, LK, LR, LS, LT, LU, LV, MA, MD, MG, MK, MN, MW, MX, MZ, NO, NZ, PL, PT, RO, RU, SD, SE, SG, SI, SK, SL, TJ, TM, TR, TT, TZ, UA, UG, UZ, VN, YU, ZA, ZW.

(84) Designated States *(regional)*: ARIPO patent (GH, GM, KE, LS, MW, MZ, SD, SL, SZ, TZ, UG, ZW), Eurasian patent (AM, AZ, BY, KG, KZ, MD, RU, TJ, TM), European patent (AT, BE, CH, CY, DE, DK, ES, FI, FR, GB, GR, IE, IT, LU, MC, NL, PT, SE, TR), OAPI patent (BF, BJ, CF, CG, CI, CM, GA, GN, GW, ML, MR, NE, SN, TD, TG).

Published:
— *with international search report*

For two-letter codes and other abbreviations, refer to the "Guidance Notes on Codes and Abbreviations" appearing at the beginning of each regular issue of the PCT Gazette.

WO 02/30981 A1

(54) Title: GENE ENCODING A HUMAN G-PROTEIN COUPLED RECEPTOR AND ITS USE

(57) Abstract: The present invention provides amino acid sequences of peptides that are encoded by genes within the Human genome, the GPCR peptides of the present invention. The present invention specifically provides isolated peptide and nucleic acid molecules, methods of identifying orthologs and paralogs of the GPCR peptides and methods of identifying modulators of the GPCR peptides.

GENE ENCODING A HUMAN G-PROTEIN COUPLED RECEPTOR AND ITS USE

RELATED APPLICATIONS

The present application claims priority to U.S. Serial No. 09/684,393, filed October 10, 2000 (Atty. Docket CL000869).

FIELD OF THE INVENTION

The present invention is in the field of G-Protein coupled receptors (GPCRs) that are related to the calcium-sensing receptor subfamily, recombinant DNA molecules, and protein production. The present invention specifically provides novel GPCR peptides and proteins and nucleic acid molecules encoding such peptide and protein molecules, all of which are useful in the development of human therapeutics and diagnostic compositions and methods.

BACKGROUND OF THE INVENTION

G-protein coupled receptors

G-protein coupled receptors (GPCRs) constitute a major class of proteins responsible for transducing a signal within a cell. GPCRs have three structural domains: an amino terminal extracellular domain, a transmembrane domain containing seven transmembrane segments, three extracellular loops, and three intracellular loops, and a carboxy terminal intracellular domain. Upon binding of a ligand to an extracellular portion of a GPCR, a signal is transduced within the cell that results in a change in a biological or physiological property of the cell. GPCRs, along with G-proteins and effectors (intracellular enzymes and channels modulated by G-proteins), are the components of a modular signaling system that connects the state of intracellular second messengers to extracellular inputs.

GPCR genes and gene-products are potential causative agents of disease (Spiegel *et al.*, *J. Clin. Invest. 92*:1119-1125 (1993); McKusick *et al.*, *J. Med. Genet. 30*:1-26 (1993)). Specific defects in the rhodopsin gene and the V2 vasopressin receptor gene have been shown to cause various forms of retinitis pigmentosum (Nathans *et al.*, *Annu. Rev. Genet. 26*:403-424(1992)), and nephrogenic diabetes insipidus (Holtzman *et al.*, *Hum. Mol. Genet. 2*:1201-1204 (1993)). These receptors are of critical importance to both the central nervous system and peripheral

physiological processes. Evolutionary analyses suggest that the ancestor of these proteins originally developed in concert with complex body plans and nervous systems.

The GPCR protein superfamily can be divided into five families: Family I, receptors typified by rhodopsin and the β2-purinergic receptor and currently represented by over 200 unique members (Dohlman *et al.*, *Annu. Rev. Biochem. 60*:653-688 (1991)); Family II, the parathyroid hormone/calcitonin/secretin receptor family (Juppner *et al.*, *Science 254*:1024-1026 (1991); Lin *et al.*, *Science 254*:1022-1024 (1991)); Family III, the metabotropic glutamate receptor family (Nakanishi, *Science 258* 597:603 (1992)); Family IV, the cAMP receptor family, important in the chemotaxis and development of *D. discoideum* (Klein *et al.*, *Science 241*:1467-1472 (1988)); and Family V, the fungal mating pheromone receptors such as STE2 (Kurjan, *Annu. Rev. Biochem. 61*:1097-1129 (1992)).

There are also a small number of other proteins that present seven putative hydrophobic segments and appear to be unrelated to GPCRs; they have not been shown to couple to G-proteins. *Drosophila* expresses a photoreceptor-specific protein, bride of sevenless (boss), a seven-transmembrane-segment protein that has been extensively studied and does not show evidence of being a GPCR (Hart *et al.*, *Proc. Natl. Acad. Sci. USA 90*:5047-5051 (1993)). The gene *frizzled* (*fz*) in *Drosophila* is also thought to be a protein with seven transmembrane segments. Like boss, fz has not been shown to couple to G-proteins (Vinson *et al.*, *Nature 338*:263-264 (1989)).

G proteins represent a family of heterotrimeric proteins composed of α, β and γ subunits, that bind guanine nucleotides. These proteins are usually linked to cell surface receptors, e.g., receptors containing seven transmembrane segments. Following ligand binding to the GPCR, a conformational change is transmitted to the G protein, which causes the α-subunit to exchange a bound GDP molecule for a GTP molecule and to dissociate from the βγ-subunits. The GTP-bound form of the α-subunit typically functions as an effector-modulating moiety, leading to the production of second messengers, such as cAMP (e.g., by activation of adenyl cyclase), diacylglycerol or inositol phosphates. Greater than 20 different types of α-subunits are known in humans. These subunits associate with a smaller pool of β and γ subunits. Examples of mammalian G proteins include Gi, Go, Gq, Gs and Gt. G proteins are described extensively in Lodish *et al.*, *Molecular Cell Biology*, (Scientific American Books Inc., New York, N.Y., 1995), the contents of which are incorporated herein by reference. GPCRs, G proteins and G protein-linked effector and second messenger systems have been reviewed in *The G-Protein Linked Receptor Fact Book*, Watson *et al.*, eds., Academic Press (1994).

Calcium-Sensing Receptors

The protein provided by the present invention is highly homologous to calcium-sensing receptors (CaRs), which are GPCRs. CaRs share extensive sequence similarity with odorant and taste receptors. Both CaRs and odorant receptors may be expressed in epithelia. CaRs form dimers held together by disulfide links. Intermolecular interactions between monomers are thought to be essential for CaR activity.

Calcium is vital to a wide array of physiological processes and therefore it is critical that the concentration of calcium in extracellular fluids be kept within a narrow range. Mutations in CaR that increase or decrease the responsiveness of the CaR receptor to extracellular calcium concentrations are associated with inherited genetic disorders of calcium homeostasis. Therefore, it is likely that CaR is the main regulator of divalent mineral ion excretion.

CaRs are expressed in, and stimulate proliferation of, fibroblasts, where they are involved in calcium-dependent activation of Src and mitogen-activated kinases in response to extracellular calcium. CaRs may also be expressed in thyroid glands and are likely involved in the etiology of hyper- and hypocalcemic disorders. Naturally occurring mutations of CaRs are associated with several inherited conditions, including familial hypocalciuric hypercalcemia and neonatal severe hyperparathyroidism.

CaRs are also involved in epithelial differentiation and CaRs may be expressed in keratinocytes where they likely play an essential role in keratinocyte division and differentiation. Deletion of CaR in knockout mice results in visible alterations of epidermis and reduced levels of loricin, a keratinocyte differentiation marker. Epidermis regeneration is a continuous process that is essential for replacement of skin as well as inner linings of organs such as intestines, kidney ducts, and thyroid. The speed of this process is under tight control of regulatory factors, many of which are unknown. It is possible that CaR levels are elevated in rapidly dividing skin cells, for example, in keratomas and breast tumors. Antibodies derived against CaRs may be used to detect tumors, and synthetic peptide inhibitors that bind CaRs and block its ability to detect calcium may be used as anti-cancer drugs. Short peptides that mimic the CaR dimerization domain could prevent assembly of functional CaRs.

For a further review of CaRs, see: Oda *et al.*, *J Biol Chem* 2000 Jan 14;275(2):1183-90; Bikle et al., *J Clin Invest* 1996 Feb 15;97(4):1085-93; McNeil *et al.*, *J Biol Chem* 1998 Jan 9;273(2):1114-20; Bai *et al.*, *Proc Natl Acad Sci U S A* 1999 Mar 16;96(6):2834-9;

Emanuel *et al.*, *Mol Endocrinol* 1996 May;10(5):555-65; and Riccardi *et al.*, *Arch Med Res* 1999 Nov-Dec,30(6):436-48.

Taste Receptors

Two GPCRs, which have been identified in the apical membranes of rat and mouse taste cells and are differentially dispersed on the tongue and palate, are putative taste receptors. These receptors are targets for studies into gustatory processing. These putative taste receptors show extensive sequence similarity to calcium-sensing receptors.

For a further review of putative taste receptors, see Smith *et al.*, *Curr Biol* 1999 Jun 17, 9(12): R453-5.

Aminergic GPCRs

One family of the GPCRS, Family II, contains receptors for acetylcholine, catecholamine, and indoleamine ligands (hereafter referred to as biogenic amines). The biogenic amine receptors (aminergic GPCRs) represent a large group of GPCRs that share a common evolutionary ancestor and which are present in both vertebrate (deuterostome), and invertebrate (protostome) lineages. This family of GPCRs includes, but is not limited to the 5-HT-like, the dopamine-like, the acetylcholine-like, the adrenaline-like and the melatonin-like GPCRs.

Dopamine receptors

The understanding of the dopaminergic system relevance in brain function and disease developed several decades ago from three diverse observations following drug treatments. These were the observations that dopamine replacement therapy improved Parkinson's disease symptoms, depletion of dopamine and other catecholamines by reserpine caused depression and antipsychotic drugs blocked dopamine receptors. The finding that the dopamine receptor binding affinities of typical antipsychotic drugs correlate with their clinical potency led to the dopamine overactivity hypothesis of schizophrenia (Snyder, S.H., *Am J Psychiatry 133*, 197-202 (1976); Seeman, P. and Lee, T., *Science 188*, 1217-9 (1975)). Today, dopamine receptors are crucial targets in the pharmacological therapy of schizophrenia, Parkinson's disease, Tourette's syndrome, tardive dyskinesia and Huntington's disease. The dopaminergic system includes the nigrostriatal, mesocorticolimbic and tuberoinfundibular pathways. The nigrostriatal pathway is part of the striatal motor system and its degeneration leads to Parkinson's disease; the mesocorticolimbic pathway plays a key role in reinforcement and in emotional expression and is

the desired site of action of antipsychotic drugs; the tuberoinfundibular pathways regulates prolactin secretion from the pituitary.

Dopamine receptors are members of the G protein coupled receptor superfamily, a large group proteins that share a seven helical membrane-spanning structure and transduce signals through coupling to heterotrimeric guanine nucleotide-binding regulatory proteins (G proteins). Dopamine receptors are classified into subfamilies: D1-like (D1 and D5) and D2-like (D2, D3 and D4) based on their different ligand binding profiles, signal transduction properties, sequence homologies and genomic organizations (Civelli, O., Bunzow, J.R. and Grandy, D.K., *Annu Rev Pharmacol Toxicol 33*, 281-307 (1993)). The D1-like receptors, D1 and D5, stimulate cAMP synthesis through coupling with Gs-like proteins and their genes do not contain introns within their protein coding regions. On the other hand, the D2-like receptors, D2, D3 and D4, inhibit cAMP synthesis through their interaction with Gi-like proteins and share a similar genomic organization which includes introns within their protein coding regions.

Serotonin receptors

Serotonin (5-Hydroxytryptamine; 5-HT) was first isolated from blood serum, where it was shown to promote vasoconstriction (Rapport, M.M., Green, A.A. and Page, I.H., *J Biol Chem 176*, 1243-1251 (1948). Interest on a possible relationship between 5-HT and psychiatric disease was spurred by the observations that hallucinogens such as LSD and psilocybin inhibit the actions of 5-HT on smooth muscle preparations (Gaddum, J.H. and Hameed, K.A., *Br J Pharmacol 9*, 240-248 (1954)). This observation lead to the hypothesis that brain 5-HT activity might be altered in psychiatric disorders (Wooley, D.W. and Shaw, E., *Proc Natl Acad Sci U S A 40*, 228-231 (1954); Gaddum, J.H. and Picarelli, Z.P., *Br J Pharmacol 12*, 323-328 (1957)). This hypothesis was strengthened by the introduction of tricyclic antidepressants and monoamine oxidase inhibitors for the treatment of major depression and the observation that those drugs affected noradrenaline and 5-HT metabolism. Today, drugs acting on the serotoninergic system have been proved to be effective in the pharmacotherapy of psychiatric diseases such as depression, schizophrenia, obsessive-compulsive disorder, panic disorder, generalized anxiety disorder and social phobia as well as migraine, vomiting induced by cancer chemotherapy and gastric motility disorders.

Serotonin receptors represent a very large and diverse family of neurotransmitter receptors. To date thirteen 5-HT receptor proteins coupled to G proteins plus one ligand–gated ion channel receptor (5-HT3) have been described in mammals. This receptor diversity is thought to reflect serotonin's ancient origin as a neurotransmitter and a hormone as well as the

many different roles of 5-HT in mammals. The 5-HT receptors have been classified into seven subfamilies or groups according to their different ligand-binding affinity profiles, molecular structure and intracellular transduction mechanisms (Hoyer, D. et al., *Pharmacol. Rev. 46*, 157-203 (1994)).

Adrenergic GPCRs

The adrenergic receptors comprise one of the largest and most extensively characterized families within the G-protein coupled receptor "superfamily". This superfamily includes not only adrenergic receptors, but also muscarinic, cholinergic, dopaminergic, serotonergic, and histaminergic receptors. Numerous peptide receptors include glucagon, somatostatin, and vasopressin receptors, as well as sensory receptors for vision (rhodopsin), taste, and olfaction, also belong to this growing family. Despite the diversity of signalling molecules, G-protein coupled receptors all possess a similar overall primary structure, characterized by 7 putative membrane-spanning .alpha. helices (Probst et al., 1992). In the most basic sense, the adrenergic receptors are the physiological sites of action of the catecholamines, epinephrine and norepinephrine. Adrenergic receptors were initially classified as either .alpha. or .beta. by Ahlquist, who demonstrated that the order of potency for a series of agonists to evoke a physiological response was distinctly different at the 2 receptor subtypes (Ahlquist, 1948). Functionally, .alpha. adrenergic receptors were shown to control vasoconstriction, pupil dilation and uterine inhibition, while .beta. adrenergic receptors were implicated in vasorelaxation, myocardial stimulation and bronchodilation (Regan et al., 1990). Eventually, pharmacologists realized that these responses resulted from activation of several distinct adrenergic receptor subtypes. .beta. adrenergic receptors in the heart were defined as .beta..sub.1, while those in the lung and vasculature were termed .beta..sub.2 (Lands et al., 1967).

.alpha. Adrenergic receptors, meanwhile, were first classified based on their anatomical location, as either pre or post-synaptic (.alpha..sub.2 and .alpha..sub.1, respectively) (Langer et al., 1974). This classification scheme was confounded, however, by the presence of .alpha..sub.2 receptors in distinctly non-synaptic locations, such as platelets (Berthelsen and Pettinger, 1977). With the development of radioligand binding techniques, .alpha. adrenergic receptors could be distinguished pharmacologically based on their affinities for the antagonists prazosin or yohimbine (Stark, 1981). Definitive evidence for adrenergic receptor subtypes, however, awaited purification and molecular cloning of adrenergic receptor subtypes. In 1986, the genes for the hamster .beta..sub.2 (Dickson et al.,

1986) and turkey .beta..sub.1 adrenergic receptors (Yarden et al., 1986) were cloned and sequenced. Hydropathy analysis revealed that these proteins contain 7 hydrophobic domains similar to rhodopsin, the receptor for light. Since that time the adrenergic receptor family has expanded to include 3 subtypes of .beta. receptors (Emorine et al., 1989), 3 subtypes of .alpha..sub.1 receptors (Schwinn et al., 1990), and 3 distinct types of .beta..sub.2 receptors (Lomasney et al., 1990).

The cloning, sequencing and expression of alpha receptor subtypes from animal tissues has led to the subclassification of the alpha 1 receptors into alpha 1d (formerly known as alpha 1a or 1a/1d), alpha 1b and alpha 1a (formerly known as alpha 1c) subtypes. Each alpha 1 receptor subtype exhibits its own pharmacologic and tissue specificities. The designation "alpha 1a" is the appellation recently approved by the IUPHAR Nomenclature Committee for the previously designated "alpha 1c" cloned subtype as outlined in the 1995 Receptor and Ion Channel Nomenclature Supplement (Watson and Girdlestone, 1995). The designation alpha 1a is used throughout this application to refer to this subtype. At the same time, the receptor formerly designated alpha 1a was renamed alpha 1d. The new nomenclature is used throughout this application. Stable cell lines expressing these alpha 1 receptor subtypes are referred to herein; however, these cell lines were deposited with the American Type Culture Collection (ATCC) under the old nomenclature. For a review of the classification of alpha 1 adrenoceptor subtypes, see, Martin C. Michel, et al., Naunyn-Schmiedeberg's Arch. Pharmacol. (1995) 352:1-10.

The differences in the alpha adrenergic receptor subtypes have relevance in pathophysiologic conditions. Benign prostatic hyperplasia, also known as benign prostatic hypertrophy or BPH, is an illness typically affecting men over fifty years of age, increasing in severity with increasing age. The symptoms of the condition include, but are not limited to, increased difficulty in urination and sexual dysfunction. These symptoms are induced by enlargement, or hyperplasia, of the prostate gland. As the prostate increases in size, it impinges on free-flow of fluids through the male urethra. Concommitantly, the increased noradrenergic innervation of the enlarged prostate leads to an increased adrenergic tone of the bladder neck and urethra, further restricting the flow of urine through the urethra.

The .alpha..sub.2 receptors appear to have diverged rather early from either .beta. or .alpha..sub.1 receptors. The .alpha..sub.2 receptors have been broken down into 3 molecularly distinct subtypes termed .alpha..sub.2 C2, .alpha..sub.2 C4, and .alpha..sub.2 C10 based on their chromosomal location. These subtypes appear to correspond to the pharmacologically defined .alpha..sub.2B, .alpha..sub.2C, and .alpha..sub.2A subtypes, respectively (Bylund et

al., 1992). While all the receptors of the adrenergic type are recognized by epinephrine, they are pharmacologically distinct and are encoded by separate genes. These receptors are generally coupled to different second messenger pathways that are linked through G-proteins. Among the adrenergic receptors, .beta..sub.1 and .beta..sub.2 receptors activate the adenylate cyclase, .alpha..sub.2 receptors inhibit adenylate cyclase and .alpha..sub.1 receptors activate phospholipase C pathways, stimulating breakdown of polyphosphoinositides (Chung, F. Z. et al., J. Biol. Chem., 263:4052 (1988)). .alpha..sub.1 and .alpha..sub.2 adrenergic receptors differ in their cell activity for drugs.

Issued US patent that disclose the utility of members of this family of proteins include, but are not limited to, 6,063,785 Phthalimido arylpiperazines useful in the treatment of benign prostatic hyperplasia; 6,060,492 Selective .beta.3 adrenergic agonists; 6,057,350 Alpha 1a adrenergic receptor antagonists; 6,046,192 Phenylethanolaminotetralincarboxamide derivatives; 6,046,183 Method of synergistic treatment for benign prostatic hyperplasia; 6,043,253 Fused piperidine substituted arylsulfonamides as .beta.3-agonists; 6,043,224 Compositions and methods for treatment of neurological disorders and neurodegenerative diseases; 6,037,354 Alpha 1a adrenergic receptor antagonists; 6,034,106 Oxadiazole benzenesulfonamides as selective .beta..sub.3 Agonist for the treatment of Diabetes and Obesity; 6,011,048 Thiazole benzenesulfonamides as .beta.3 agonists for treatment of diabetes and obesity; 6,008,361 5,994,506 Adrenergic receptor; 5,994,294 Nitrosated and nitrosylated .alpha.-adrenergic receptor antagonist compounds, compositions and their uses; 5,990,128 .alpha..sub.1C specific compounds to treat benign prostatic hyperplasia; 5,977,154 Selective .beta.3 adrenergic agonist; 5,977,115 Alpha 1a adrenergic receptor antagonists; 5,939,443 Selective .beta.3 adrenergic agonists; 5,932,538 Nitrosated and nitrosylated .alpha.-adrenergic receptor antagonist compounds, compositions and their uses; 5,922,722 Alpha 1a adrenergic receptor antagonists 26 5,908,830 and 5,861,309 DNA endoding human alpha 1 adrenergic receptors.

Purinergic GPCRs

Purinoceptor P2Y1

P2 purinoceptors have been broadly classified as P2X receptors which are ATP-gated channels; P2Y receptors, a family of G protein-coupled receptors, and P2Z receptors, which mediate nonselective pores in mast cells. Numerous subtypes have been identified for each of

the P2 receptor classes. P2Y receptors are characterized by their selective responsiveness towards ATP and its analogs. Some respond also to UTP. Based on the recommendation for nomenclature of P2 purinoceptors, the P2Y purinoceptors were numbered in the order of cloning. P2Y1, P2Y2 and P2Y3 have been cloned from a variety of species. P2Y1 responds to both ADP and ATP. Analysis of P2Y receptor subtype expression in human bone and 2 osteoblastic cell lines by RT-PCR showed that all known human P2Y receptor subtypes were expressed: P2Y1, P2Y2, P2Y4, P2Y6, and P2Y7 (Maier et al. 1997). In contrast, analysis of brain-derived cell lines suggested that a selective expression of P2Y receptor subtypes occurs in brain tissue.

Leon et al. generated P2Y1-null mice to define the physiologic role of the P2Y1 receptor. (J. Clin. Invest. 104: 1731-1737(1999)) These mice were viable with no apparent abnormalities affecting their development, survival, reproduction, or morphology of platelets, and the platelet count in these animals was identical to that of wildtype mice. However, platelets from P2Y1-deficient mice were unable to aggregate in response to usual concentrations of ADP and displayed impaired aggregation to other agonists, while high concentrations of ADP induced platelet aggregation without shape change. In addition, ADP-induced inhibition of adenylyl cyclase still occurred, demonstrating the existence of an ADP receptor distinct from P2Y1. P2Y1-null mice had no spontaneous bleeding tendency but were resistant to thromboembolism induced by intravenous injection of ADP or collagen and adrenaline. Hence, the P2Y1 receptor plays an essential role in thrombotic states and represents a potential target for antithrombotic drugs. Somers et al. mapped the P2RY1 gene between flanking markers D3S1279 and D3S1280 at a position 173 to 174 cM from the most telomeric markers on the short arm of chromosome 3. (Genomics 44: 127-130 (1997)).

Purinoceptor P2Y2

The chloride ion secretory pathway that is defective in cystic fibrosis (CF) can be bypassed by an alternative pathway for chloride ion transport that is activated by extracellular nucleotides. Accordingly, the P2 receptor that mediates this effect is a therapeutic target for improving chloride secretion in CF patients. Parr et al. reported the sequence and functional expression of a cDNA cloned from human airway epithelial cells that encodes a protein with properties of a P2Y nucleotide receptor. (Proc. Nat. Acad. Sci. 91: 3275-3279 (1994)) The human P2RY2 gene was mapped to chromosome 11q13.5-q14.1.

Purinoceptor P2RY4

The P2RY4 receptor appears to be activated specifically by UTP and UDP, but not by ATP and ADP. Activation of this uridine nucleotide receptor resulted in increased inositol phosphate formation and calcium mobilization. The UNR gene is located on chromosome Xq13.

Purinoceptor P2Y6

Somers et al. mapped the P2RY6 gene to 11q13.5, between polymorphic markers D11S1314 and D11S916, and P2RY2 maps within less than 4 cM of P2RY6. (Genomics 44: 127-130 (1997)) This was the first chromosomal clustering of this gene family to be described.

Adenine and uridine nucleotides, in addition to their well established role in intracellular energy metabolism, phosphorylation, and nucleic acid synthesis, also are important extracellular signaling molecules. P2Y metabotropic receptors are GPCRs that mediate the effects of extracellular nucleotides to regulate a wide variety of physiological processes. At least ten subfamilies of P2Y receptors have been identified. These receptor subfamilies differ greatly in their sequences and in their nucleotide agonist selectivities and efficacies.

It has been demonstrated that the P2Y1 receptors are strongly expressed in the brain, but the P2Y2, P2Y4 and P2Y6 receptors are also present. The localisation of one or more of these subtypes on neurons, on glia cells, on brain vasculature or on ventricle ependimal cells was found by in situ mRNA hybridisation and studies on those cells in culture. The P2Y1 receptors are prominent on neurons. The coupling of certain P2Y receptor subtypes to N-type Ca2+ channels or to particular K+ channels was also demonstrated.

It has also been demonstrated that several P2Y receptors mediate potent growth stimulatory effects on smooth muscle cells by stimulating intracellular pathways including Gq-proteins, protein kinase C and tyrosine phosphorylation, leading to increased immediate early gene expression, cell number, DNA and protein synthesis. It has been further demonstrated that P2Y regulation plays a mitogenic role in response to the development of artherosclerosis.

It has further been demonstrated that P2Y receptors play a critical role in cystic fibrosis. The volume and composition of the liquid that lines the airway surface is modulated by active transport of ions across the airway epithelium. This in turn is regulated both by autonomic agonists acting on basolateral receptors and by agonists acting on luminal receptors. Specifically, extracellular nucleotides present in the airway surface liquid act on luminal P2Y receptors to control both Cl- secretion and Na+ absorption. Since nucleotides are released in a regulated manner from airway epithelial cells, it is likely that their control over airway ion

transport forms part of an autocrine regulatory system localised to the luminal surface of airway epithelia. In addition to this physiological role, P2Y receptor agonists have the potential to be of crucial benefit in the treatment of CF, a disorder of epithelial ion transport. The airways of people with CF have defective Cl- secretion and abnormally high rates of Na+ absorption. Since P2Y receptor agonists can regulate both these ion transport pathways they have the potential to pharmacologically bypass the ion transport defects in CF.

GPCRs, particularly members of the calcium-sensing receptor subfamily, are a major target for drug action and development. Accordingly, it is valuable to the field of pharmaceutical development to identify and characterize previously unknown GPCRs. The present invention advances the state of the art by providing a previously unidentified human GPCR.

SUMMARY OF THE INVENTION

The present invention is based in part on the identification of nucleic acid sequences that encode amino acid sequences of human GPCR peptides and proteins that are related to the calcium-sensing receptor subfamily, allelic variants thereof and other mammalian orthologs thereof. These unique peptide sequences, and nucleic acid sequences that encode these peptides, can be used as models for the development of human therapeutic targets, aid in the identification of therapeutic proteins, and serve as targets for the development of human therapeutic agents.

The proteins of the present inventions are GPCRs that participate in signaling pathways mediated by the calcium-sensing receptor subfamily in cells that express these proteins. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. As used herein, a "signaling pathway" refers to the modulation (e.g., stimulation or inhibition) of a cellular function/activity upon the binding of a ligand to the GPCR protein. Examples of such functions include mobilization of intracellular molecules that participate in a signal transduction pathway, e.g., phosphatidylinositol 4,5-bisphosphate (PIP_2), inositol 1,4,5-triphosphate (IP_3) and adenylate cyclase; polarization of the plasma membrane; production or secretion of molecules; alteration in the structure of a cellular component; cell proliferation, e.g., synthesis of DNA; cell migration; cell differentiation; and cell survival

The response mediated by the receptor protein depends on the type of cell it is expressed on. Some information regarding the types of cells that express other members of the subfamily

of GPCRs of the present invention is already known in the art (see references cited in Background and information regarding closest homologous protein provided in Figure 2; Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells.). For example, in some cells, binding of a ligand to the receptor protein may stimulate an activity such as release of compounds, gating of a channel, cellular adhesion, migration, differentiation, etc., through phosphatidylinositol or cyclic AMP metabolism and turnover while in other cells, the binding of the ligand will produce a different result. Regardless of the cellular activity/response modulated by the particular GPCR of the present invention, a skilled artisan will clearly know that the receptor protein is a GPCR and interacts with G proteins to produce one or more secondary signals, in a variety of intracellular signal transduction pathways, e.g., through phosphatidylinositol or cyclic AMP metabolism and turnover, in a cell thus participating in a biological process in the cells or tissues that express the GPCR. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow.

As used herein, "phosphatidylinositol turnover and metabolism" refers to the molecules involved in the turnover and metabolism of phosphatidylinositol 4,5-bisphosphate (PIP_2) as well as to the activities of these molecules. PIP_2 is a phospholipid found in the cytosolic leaflet of the plasma membrane. Binding of ligand to the receptor activates, in some cells, the plasma-membrane enzyme phospholipase C that in turn can hydrolyze PIP_2 to produce 1,2-diacylglycerol (DAG) and inositol 1,4,5-triphosphate (IP_3). Once formed IP_3 can diffuse to the endoplasmic reticulum surface where it can bind an IP_3 receptor, e.g., a calcium channel protein containing an IP_3 binding site. IP_3 binding can induce opening of the channel, allowing calcium ions to be released into the cytoplasm. IP_3 can also be phosphorylated by a specific kinase to form inositol 1,3,4,5-tetraphosphate (IP_4), a molecule that can cause calcium entry into the cytoplasm from the extracellular medium. IP_3 and IP_4 can subsequently be hydrolyzed very rapidly to the inactive products inositol 1,4-biphosphate (IP_2) and inositol 1,3,4-triphosphate, respectively. These inactive products can be recycled by the cell to synthesize PIP_2. The other second messenger produced by the hydrolysis of PIP_2, namely 1,2-diacylglycerol (DAG), remains in the cell membrane where it can serve to activate the enzyme protein kinase C. Protein kinase C is usually found soluble in the cytoplasm of the cell, but upon an increase in the intracellular calcium concentration, this enzyme can move to the plasma membrane where it can

be activated by DAG. The activation of protein kinase C in different cells results in various cellular responses such as the phosphorylation of glycogen synthase, or the phosphorylation of various transcription factors, e.g., NF-kB. The language "phosphatidylinositol activity", as used herein, refers to an activity of PIP_2 or one of its metabolites.

Another signaling pathway in which the receptor may participate is the cAMP turnover pathway. As used herein, "cyclic AMP turnover and metabolism" refers to the molecules involved in the turnover and metabolism of cyclic AMP (cAMP) as well as to the activities of these molecules. Cyclic AMP is a second messenger produced in response to ligand-induced stimulation of certain G protein coupled receptors. In the cAMP signaling pathway, binding of a ligand to a GPCR can lead to the activation of the enzyme adenyl cyclase, which catalyzes the synthesis of cAMP. The newly synthesized cAMP can in turn activate a cAMP-dependent protein kinase. This activated kinase can phosphorylate a voltage-gated potassium channel protein, or an associated protein, and lead to the inability of the potassium channel to open during an action potential. The inability of the potassium channel to open results in a decrease in the outward flow of potassium, which normally repolarizes the membrane of a neuron, leading to prolonged membrane depolarization.

By targeting an agent to modulate a GPCR, the signaling activity and biological process mediated by the receptor can be agonized or antagonized in specific cells and tissues. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Such agonism and antagonism serves as a basis for modulating a biological activity in a therapeutic context (mammalian therapy) or toxic context (anti-cell therapy, e.g. anti-cancer agent).

DESCRIPTION OF THE FIGURE SHEETS

FIGURE 1 provides the nucleotide sequence of a cDNA molecule, with 5' and 3' UTR regions, which encodes the GPCR of the present invention. (SEQ ID NO:1) In addition, structure and functional information is provided, such as ATG start, stop and tissue distribution, where available, that allows one to readily determine specific uses of inventions based on this molecular sequence. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells.

FIGURE 2 provides the predicted amino acid sequence of the GPCR of the present invention. (SEQ ID NO:2) In addition structure and functional information such as protein

family, function, and modification sites is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence.

FIGURE 3 provides genomic sequences that span the gene encoding the GPCR protein of the present invention. (SEQ ID NO:3) In addition structure and functional information, such as intron/exon structure, promoter location, etc., is provided where available, allowing one to readily determine specific uses of inventions based on this molecular sequence. As illustrated in Figure 3, known SNP variations include T406C, T852C, G897A, C1433T, T5845C, and G7028A.

DETAILED DESCRIPTION OF THE INVENTION

General Description

The present invention is based on the sequencing of the human genome. During the sequencing and assembly of the human genome, analysis of the sequence information revealed previously unidentified fragments of the human genome that encode peptides that share structural and/or sequence homology to protein/peptide/domains identified and characterized within the art as being a GPCR protein or part of a GPCR protein, that are related to the calcium-sensing receptor subfamily. Utilizing these sequences, additional genomic sequences were assembled and transcript and/or cDNA sequences were isolated and characterized. Based on this analysis, the present invention provides amino acid sequences of human GPCR peptides and proteins that are related to the calcium-sensing receptor subfamily, nucleic acid sequences in the form of transcript sequences, cDNA sequences and/or genomic sequences that encode these GPCR peptides and proteins, nucleic acid variation (allelic information), tissue distribution of expression, and information about the closest art known protein/peptide/domain that has structural or sequence homology to the GPCR of the present invention.

In addition to being previously unknown, the peptides that are provided in the present invention are selected based on their ability to be used for the development of commercially important products and services. Specifically, the present peptides are selected based on homology and/or structural relatedness to known GPCR proteins of the calcium-sensing receptor subfamily and the expression pattern observed. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. The art has clearly established the commercial importance of members of

this family of proteins and proteins that have expression patterns similar to that of the present gene. Some of the more specific features of the peptides of the present invention, and the uses thereof, are described herein, particularly in the Background of the Invention and in the annotation provided in the Figures, and/or are known within the art for each of the known calcium-sensing receptor family or subfamily of GPCR proteins.

Specific Embodiments

Peptide Molecules

The present invention provides nucleic acid sequences that encode protein molecules that have been identified as being members of the GPCR family of proteins and are related to the calcium-sensing receptor subfamily (protein sequences are provided in Figure 2, transcript/cDNA sequences are provided in Figure 1 and genomic sequences are provided in Figure 3). The peptide sequences provided in Figure 2, as well as the obvious variants described herein, particularly allelic variants as identified herein and using the information in Figure 3, will be referred herein as the GPCR peptides of the present invention, GPCR peptides, or peptides/proteins of the present invention.

The present invention provides isolated peptide and protein molecules that consist of, consist essentially of, or comprise the amino acid sequences of the GPCR peptides disclosed in Figure 2, (encoded by the nucleic acid molecule shown in Figure 1, transcript/cDNA sequence, or Figure 3, genomic sequence), as well as all obvious variants of these peptides that are within the art to make and use. Some of these variants are described in detail below.

As used herein, a peptide is said to be "isolated" or "purified" when it is substantially free of cellular material or free of chemical precursors or other chemicals. The peptides of the present invention can be purified to homogeneity or other degrees of purity. The level of purification will be based on the intended use. The critical feature is that the preparation allows for the desired function of the peptide, even if in the presence of considerable amounts of other components (the features of an isolated nucleic acid molecule is discussed below).

In some uses, "substantially free of cellular material" includes preparations of the peptide having less than about 30% (by dry weight) other proteins (i.e., contaminating protein), less than about 20% other proteins, less than about 10% other proteins, or less than about 5% other proteins. When the peptide is recombinantly produced, it can also be substantially free of culture

medium, i.e., culture medium represents less than about 20% of the volume of the protein preparation.

The language "substantially free of chemical precursors or other chemicals" includes preparations of the peptide in which it is separated from chemical precursors or other chemicals that are involved in its synthesis. In one embodiment, the language "substantially free of chemical precursors or other chemicals" includes preparations of the GPCR peptide having less than about 30% (by dry weight) chemical precursors or other chemicals, less than about 20% chemical precursors or other chemicals, less than about 10% chemical precursors or other chemicals, or less than about 5% chemical precursors or other chemicals.

The isolated GPCR peptide can be purified from cells that naturally express it, purified from cells that have been altered to express it (recombinant), or synthesized using known protein synthesis methods. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. For example, a nucleic acid molecule encoding the GPCR peptide is cloned into an expression vector, the expression vector introduced into a host cell and the protein expressed in the host cell. The protein can then be isolated from the cells by an appropriate purification scheme using standard protein purification techniques. Many of these techniques are described in detail below.

Accordingly, the present invention provides proteins that consist of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). The amino acid sequence of such a protein is provided in Figure 2. A protein consists of an amino acid sequence when the amino acid sequence is the final amino acid sequence of the protein.

The present invention further provides proteins that consist essentially of the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein consists essentially of an amino acid sequence when such an amino acid sequence is present with only a few additional amino acid residues, for example from about 1 to about 100 or so additional residues, typically from 1 to about 20 additional residues in the final protein.

The present invention further provides proteins that comprise the amino acid sequences provided in Figure 2 (SEQ ID NO:2), for example, proteins encoded by the transcript/cDNA nucleic acid sequences shown in Figure 1 (SEQ ID NO:1) and the genomic sequences provided in Figure 3 (SEQ ID NO:3). A protein comprises an amino acid sequence when the amino acid

sequence is at least part of the final amino acid sequence of the protein. In such a fashion, the protein can be only the peptide or have additional amino acid molecules, such as amino acid residues (contiguous encoded sequence) that are naturally associated with it or heterologous amino acid residues/peptide sequences. Such a protein can have a few additional amino acid residues or can comprise several hundred or more additional amino acids. The preferred classes of proteins that are comprised of the GPCR peptides of the present invention are the naturally occurring mature proteins. A brief description of how various types of these proteins can be made/isolated is provided below.

The GPCR peptides of the present invention can be attached to heterologous sequences to form chimeric or fusion proteins. Such chimeric and fusion proteins comprise a GPCR peptide operatively linked to a heterologous protein having an amino acid sequence not substantially homologous to the GPCR peptide. "Operatively linked" indicates that the GPCR peptide and the heterologous protein are fused in-frame. The heterologous protein can be fused to the N-terminus or C-terminus of the GPCR peptide.

In some uses, the fusion protein does not affect the activity of the GPCR peptide *per se*. For example, the fusion protein can include, but is not limited to, enzymatic fusion proteins, for example beta-galactosidase fusions, yeast two-hybrid GAL fusions, poly-His fusions, MYC-tagged, HI-tagged and Ig fusions. Such fusion proteins, particularly poly-His fusions, can facilitate the purification of recombinant GPCR peptide. In certain host cells (e.g., mammalian host cells), expression and/or secretion of a protein can be increased by using a heterologous signal sequence.

A chimeric or fusion protein can be produced by standard recombinant DNA techniques. For example, DNA fragments coding for the different protein sequences are ligated together in-frame in accordance with conventional techniques. In another embodiment, the fusion gene can be synthesized by conventional techniques including automated DNA synthesizers. Alternatively, PCR amplification of gene fragments can be carried out using anchor primers which give rise to complementary overhangs between two consecutive gene fragments which can subsequently be annealed and re-amplified to generate a chimeric gene sequence (see Ausubel *et al.*, *Current Protocols in Molecular Biology*, 1992). Moreover, many expression vectors are commercially available that already encode a fusion moiety (e.g., a GST protein). A GPCR peptide-encoding nucleic acid can be cloned into such an expression vector such that the fusion moiety is linked in-frame to the GPCR peptide.

As mentioned above, the present invention also provides and enables obvious variants of the amino acid sequence of the proteins of the present invention, such as naturally occurring

mature forms of the peptide, allelic/sequence variants of the peptides, non-naturally occurring recombinantly derived variants of the peptides, and orthologs and paralogs of the peptides. Such variants can readily be generated using art-known techniques in the fields of recombinant nucleic acid technology and protein biochemistry. It is understood, however, that variants exclude any amino acid sequences disclosed prior to the invention.

Such variants can readily be identified/made using molecular techniques and the sequence information disclosed herein. Further, such variants can readily be distinguished from other peptides based on sequence and/or structural homology to the GPCR peptides of the present invention. The degree of homology/identity present will be based primarily on whether the peptide is a functional variant or non-functional variant, the amount of divergence present in the paralog family and the evolutionary distance between the orthologs.

To determine the percent identity of two amino acid sequences or two nucleic acid sequences, the sequences are aligned for optimal comparison purposes (e.g., gaps can be introduced in one or both of a first and a second amino acid or nucleic acid sequence for optimal alignment and non-homologous sequences can be disregarded for comparison purposes). In a preferred embodiment, the length of a reference sequence aligned for comparison purposes is at least 30%, 40%, 50%, 60%, 70%, 80%, or 90% or more of the length of the reference sequence. The amino acid residues or nucleotides at corresponding amino acid positions or nucleotide positions are then compared. When a position in the first sequence is occupied by the same amino acid residue or nucleotide as the corresponding position in the second sequence, then the molecules are identical at that position (as used herein amino acid or nucleic acid "identity" is equivalent to amino acid or nucleic acid "homology"). The percent identity between the two sequences is a function of the number of identical positions shared by the sequences, taking into account the number of gaps, and the length of each gap, which need to be introduced for optimal alignment of the two sequences.

The comparison of sequences and determination of percent identity and similarity between two sequences can be accomplished using a mathematical algorithm. (*Computational Molecular Biology*, Lesk, A.M., ed., Oxford University Press, New York, 1988; *Biocomputing: Informatics and Genome Projects*, Smith, D.W., ed., Academic Press, New York, 1993; *Computer Analysis of Sequence Data, Part 1*, Griffin, A.M., and Griffin, H.G., eds., Humana Press, New Jersey, 1994; *Sequence Analysis in Molecular Biology*, von Heinje, G., Academic Press, 1987; and *Sequence Analysis Primer*, Gribskov, M. and Devereux, J., eds., M Stockton Press, New York, 1991). In a preferred embodiment, the percent identity between two amino acid sequences is determined using the Needleman and Wunsch (*J. Mol. Biol.* (48):444-453

(1970)) algorithm which has been incorporated into the GAP program in the GCG software package (available at http://www.gcg.com), using either a Blossom 62 matrix or a PAM250 matrix, and a gap weight of 16, 14, 12, 10, 8, 6, or 4 and a length weight of 1, 2, 3, 4, 5, or 6. In yet another preferred embodiment, the percent identity between two nucleotide sequences is determined using the GAP program in the GCG software package (Devereux, J., *et al., Nucleic Acids Res. 12(1)*:387 (1984)) (available at http://www.gcg.com), using a NWSgapdna.CMP matrix and a gap weight of 40, 50, 60, 70, or 80 and a length weight of 1, 2, 3, 4, 5, or 6. In another embodiment, the percent identity between two amino acid or nucleotide sequences is determined using the algorithm of E. Meyers and W. Miller (CABIOS, 4:11-17 (1989)) which has been incorporated into the ALIGN program (version 2.0), using a PAM120 weight residue table, a gap length penalty of 12 and a gap penalty of 4.

The nucleic acid and protein sequences of the present invention can further be used as a "query sequence" to perform a search against sequence databases to, for example, identify other family members or related sequences. Such searches can be performed using the NBLAST and XBLAST programs (version 2.0) of Altschul, et al. (*J. Mol. Biol.* 215:403-10 (1990)). BLAST nucleotide searches can be performed with the NBLAST program, score = 100, wordlength = 12 to obtain nucleotide sequences homologous to the nucleic acid molecules of the invention. BLAST protein searches can be performed with the XBLAST program, score = 50, wordlength = 3 to obtain amino acid sequences homologous to the proteins of the invention. To obtain gapped alignments for comparison purposes, Gapped BLAST can be utilized as described in Altschul et al. (*Nucleic Acids Res.* 25(17):3389-3402 (1997)). When utilizing BLAST and gapped BLAST programs, the default parameters of the respective programs (e.g., XBLAST and NBLAST) can be used.

Full-length pre-processed forms, as well as mature processed forms, of proteins that comprise one of the peptides of the present invention can readily be identified as having complete sequence identity to one of the GPCR peptides of the present invention as well as being encoded by the same genetic locus as the GPCR peptide provided herein. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

Allelic variants of a GPCR peptide can readily be identified as being a human protein having a high degree (significant) of sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by the same genetic locus as the GPCR peptide provided herein. Genetic locus can readily be determined based on the genomic information provided in Figure 3, such as the genomic sequence mapped to the reference human. As

indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1. As used herein, two proteins (or a region of the proteins) have significant homology when the amino acid sequences are typically at least about 70-80%, 80-90%, and more typically at least about 90-95% or more homologous. A significantly homologous amino acid sequence, according to the present invention, will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under stringent conditions as more fully described below.

Figure 3 provides information on SNPs that have been found in a gene encoding the GPCR proteins of the present invention. The following SNPs were found: T406C, T852C, G897A, C1433T, T5845C, and G7028A.

Paralogs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide, as being encoded by a gene from humans, and as having similar activity or function. Two proteins will typically be considered paralogs when the amino acid sequences are typically at least about 60% or greater, and more typically at least about 70% or greater homology through a given region or domain. Such paralogs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions as more fully described below.

Orthologs of a GPCR peptide can readily be identified as having some degree of significant sequence homology/identity to at least a portion of the GPCR peptide as well as being encoded by a gene from another organism. Preferred orthologs will be isolated from mammals, preferably primates, for the development of human therapeutic targets and agents. Such orthologs will be encoded by a nucleic acid sequence that will hybridize to a GPCR peptide encoding nucleic acid molecule under moderate to stringent conditions, as more fully described below, depending on the degree of relatedness of the two organisms yielding the proteins.

Non-naturally occurring variants of the GPCR peptides of the present invention can readily be generated using recombinant techniques. Such variants include, but are not limited to deletions, additions and substitutions in the amino acid sequence of the GPCR peptide. For example, one class of substitutions are conserved amino acid substitution. Such substitutions are those that substitute a given amino acid in a GPCR peptide by another amino acid of like characteristics. Typically seen as conservative substitutions are the replacements, one for another, among the aliphatic amino acids Ala, Val, Leu, and Ile; interchange of the hydroxyl residues Ser and Thr; exchange of the acidic residues Asp and Glu; substitution between the

amide residues Asn and Gln; exchange of the basic residues Lys and Arg; and replacements among the aromatic residues Phe and Tyr. Guidance concerning which amino acid changes are likely to be phenotypically silent are found in Bowie *et al.*, *Science 247*:1306-1310 (1990).

Variant GPCR peptides can be fully functional or can lack function in one or more activities, e.g. ability to bind ligand, ability to bind G-protein, ability to mediate signaling, etc. Fully functional variants typically contain only conservative variation or variation in non-critical residues or in non-critical regions. Figure 2 provides the result of protein analysis that identifies critical domains/regions. Functional variants can also contain substitution of similar amino acids that result in no change or an insignificant change in function. Alternatively, such substitutions may positively or negatively affect function to some degree.

Non-functional variants typically contain one or more non-conservative amino acid substitutions, deletions, insertions, inversions, or truncation or a substitution, insertion, inversion, or deletion in a critical residue or critical region.

Amino acids that are essential for function can be identified by methods known in the art, such as site-directed mutagenesis or alanine-scanning mutagenesis (Cunningham *et al.*, *Science 244*:1081-1085 (1989)), particularly using the results provided in Figure 2. The latter procedure introduces single alanine mutations at every residue in the molecule. The resulting mutant molecules are then tested for biological activity such as ligand/effector molecule binding or in assays such as an *in vitro* proliferative activity. Sites that are critical for ligand-receptor binding can also be determined by structural analysis such as crystallization, nuclear magnetic resonance or photoaffinity labeling (Smith *et al., J. Mol. Biol. 224*:899-904 (1992); de Vos *et al. Science 255*:306-312 (1992)).

The present invention further provides fragments of the GPCR peptides, in addition to proteins and peptides that comprise and consist of such fragments, particularly those comprising the residues identified in Figure 2. The fragments to which the invention pertains, however, are not to be construed as encompassing fragments that may be disclosed publicly prior to the present invention.

As used herein, a fragment comprises at least 8, 10, 12, 14, 16, or more contiguous amino acid residues from a GPCR peptide. Such fragments can be chosen based on the ability to retain one or more of the biological activities of the GPCR peptide or could be chosen for the ability to perform a function, e.g. ability to bind ligand or effector molecule or act as an immunogen. Particularly important fragments are biologically active fragments, peptides which are, for example, about 8 or more amino acids in length. Such fragments will typically comprise a domain or motif of the GPCR peptide, e.g., active site, a G-protein binding site, a

transmembrane domain or a ligand-binding domain. Further, possible fragments include, but are not limited to, domain or motif containing fragments, soluble peptide fragments, and fragments containing immunogenic structures. Predicted domains and functional sites are readily identifiable by computer programs well-known and readily available to those of skill in the art (e.g., PROSITE analysis). The results of one such analysis are provided in Figure 2.

Polypeptides often contain amino acids other than the 20 amino acids commonly referred to as the 20 naturally occurring amino acids. Further, many amino acids, including the terminal amino acids, may be modified by natural processes, such as processing and other post-translational modifications, or by chemical modification techniques well known in the art. Common modifications that occur naturally in GPCR peptides are described in basic texts, detailed monographs, and the research literature, and they are well known to those of skill in the art(some of these features are identified in Figure 2).

Known modifications include, but are not limited to, acetylation, acylation, ADP-ribosylation, amidation, covalent attachment of flavin, covalent attachment of a heme moiety, covalent attachment of a nucleotide or nucleotide derivative, covalent attachment of a lipid or lipid derivative, covalent attachment of phosphotidylinositol, cross-linking, cyclization, disulfide bond formation, demethylation, formation of covalent crosslinks, formation of cystine, formation of pyroglutamate, formylation, gamma carboxylation, glycosylation, GPI anchor formation, hydroxylation, iodination, methylation, myristoylation, oxidation, proteolytic processing, phosphorylation, prenylation, racemization, selenoylation, sulfation, transfer-RNA mediated addition of amino acids to proteins such as arginylation, and ubiquitination.

Such modifications are well-known to those of skill in the art and have been described in great detail in the scientific literature. Several particularly common modifications, glycosylation, lipid attachment, sulfation, gamma-carboxylation of glutamic acid residues, hydroxylation and ADP-ribosylation, for instance, are described in most basic texts, such as *Proteins - Structure and Molecular Properties*, 2nd Ed., T.E. Creighton, W. H. Freeman and Company, New York (1993). Many detailed reviews are available on this subject, such as by Wold, F., *Posttranslational Covalent Modification of Proteins*, B.C. Johnson, Ed., Academic Press, New York 1-12 (1983); Seifter *et al.* (*Meth. Enzymol. 182*: 626-646 (1990)) and Rattan *et al.* (*Ann. N.Y. Acad. Sci. 663*:48-62 (1992)).

Accordingly, the GPCR peptides of the present invention also encompass derivatives or analogs in which a substituted amino acid residue is not one encoded by the genetic code, in which a substituent group is included, in which the mature GPCR peptide is fused with another compound, such as a compound to increase the half-life of the GPCR peptide (for example,

polyethylene glycol), or in which the additional amino acids are fused to the mature GPCR peptide, such as a leader or secretory sequence or a sequence for purification of the mature GPCR peptide or a pro-protein sequence.

Protein/Peptide Uses

The proteins of the present invention can be used in substantial and specific assays related to the functional information provided in the Figures and Back Ground Section; to raise antibodies or to elicit another immune response; as a reagent (including the labeled reagent) in assays designed to quantitatively determine levels of the protein (or its binding partner or receptor) in biological fluids; and as markers for tissues in which the corresponding protein is preferentially expressed (either constitutively or at a particular stage of tissue differentiation or development or in a disease state). Where the protein binds or potentially binds to another protein (such as, for example, in a receptor-ligand interaction), the protein can be used to identify the binding partner so as to develop a system to identify inhibitors of the binding interaction. Any or all of these research utilities are capable of being developed into reagent grade or kit format for commercialization as commercial products.

Methods for performing the uses listed above are well known to those skilled in the art. References disclosing such methods include "Molecular Cloning: A Laboratory Manual", 2d ed., Cold Spring Harbor Laboratory Press, Sambrook, J., E. F. Fritsch and T. Maniatis eds., 1989, and "Methods in Enzymology: Guide to Molecular Cloning Techniques", Academic Press, Berger, S. L. and A. R. Kimmel eds., 1987.

The potential uses of the peptides of the present invention are based primarily on the source of the protein as well as the class/action of the protein. For example, GPCRs isolated from humans and their human/mammalian orthologs serve as targets for identifying agents for use in mammalian therapeutic applications, e.g. a human drug, particularly in modulating a biological or pathological response in a cell or tissue that expresses the GPCR. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow. Approximately 70% of all pharmaceutical agents modulate the activity of a GPCR. A combination of the invertebrate and mammalian ortholog can be used in selective screening methods to find agents specific for invertebrates.

The structural and functional information provided in the Background and Figures provide specific and substantial uses for the molecules of the present invention, particularly in combination with the expression information provided in Figure 1. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Such uses can readily be determined using the information provided herein, that known in the art and routine experimentation.

The proteins of the present invention (including variants and fragments that may have been disclosed prior to the present invention) are useful for biological assays related to GPCRs that are related to members of the calcium-sensing receptor subfamily. Such assays involve any of the known GPCR functions or activities or properties useful for diagnosis and treatment of GPCR-related conditions that are specific for the subfamily of GPCRs that the one of the present invention belongs to, particularly in cells and tissues that express this receptor. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow.

The proteins of the present invention are also useful in drug screening assays, in cell-based or cell-free systems. Cell-based systems can be native, i.e., cells that normally express the receptor protein, as a biopsy or expanded in cell culture. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. In an alternate embodiment, cell-based assays involve recombinant host cells expressing the receptor protein.

The polypeptides can be used to identify compounds that modulate receptor activity of the protein in its natural state, or an altered form that causes a specific disease or pathology associated with the receptor. Both the GPCRs of the present invention and appropriate variants and fragments can be used in high-throughput screens to assay candidate compounds for the ability to bind to the receptor. These compounds can be further screened against a functional receptor to determine the effect of the compound on the receptor activity. Further, these compounds can be tested in animal or invertebrate systems to determine activity/effectiveness. Compounds can be identified that activate (agonist) or inactivate (antagonist) the receptor to a desired degree.

Further, the proteins of the present invention can be used to screen a compound for the ability to stimulate or inhibit interaction between the receptor protein and a molecule that

normally interacts with the receptor protein, e.g. a ligand or a component of the signal pathway that the receptor protein normally interacts (for example, a G-protein or other interactor involved in cAMP or phosphatidylinositol turnover and/or adenylate cyclase, or phospholipase C activation). Such assays typically include the steps of combining the receptor protein with a candidate compound under conditions that allow the receptor protein, or fragment, to interact with the target molecule, and to detect the formation of a complex between the protein and the target or to detect the biochemical consequence of the interaction with the receptor protein and the target, such as any of the associated effects of signal transduction such as G-protein phosphorylation, cAMP or phosphatidylinositol turnover, and adenylate cyclase or phospholipase C activation.

Candidate compounds include, for example, 1) peptides such as soluble peptides, including Ig-tailed fusion peptides and members of random peptide libraries (see, e.g., Lam *et al.*, *Nature 354*:82-84 (1991); Houghten *et al.*, *Nature 354*:84-86 (1991)) and combinatorial chemistry-derived molecular libraries made of D- and/or L- configuration amino acids; 2) phosphopeptides (e.g., members of random and partially degenerate, directed phosphopeptide libraries, see, e.g., Songyang *et al.*, *Cell 72*:767-778 (1993)); 3) antibodies (e.g., polyclonal, monoclonal, humanized, anti-idiotypic, chimeric, and single chain antibodies as well as Fab, $F(ab')_2$, Fab expression library fragments, and epitope-binding fragments of antibodies); and 4) small organic and inorganic molecules (e.g., molecules obtained from combinatorial and natural product libraries).

One candidate compound is a soluble fragment of the receptor that competes for ligand binding. Other candidate compounds include mutant receptors or appropriate fragments containing mutations that affect receptor function and thus compete for ligand. Accordingly, a fragment that competes for ligand, for example with a higher affinity, or a fragment that binds ligand but does not allow release, is encompassed by the invention.

The invention further includes other end point assays to identify compounds that modulate (stimulate or inhibit) receptor activity. The assays typically involve an assay of events in the signal transduction pathway that indicate receptor activity. Thus, a cellular process such as proliferation, the expression of genes that are up- or down-regulated in response to the receptor protein dependent signal cascade, can be assayed. In one embodiment, the regulatory region of such genes can be operably linked to a marker that is easily detectable, such as luciferase.

Any of the biological or biochemical functions mediated by the receptor can be used as an endpoint assay. These include all of the biochemical or biochemical/biological events

described herein, in the references cited herein, incorporated by reference for these endpoint assay targets, and other functions known to those of ordinary skill in the art or that can be readily identified using the information provided in the Figures, particularly Figure 2. Specifically, a biological function of a cell or tissues that expresses the receptor can be assayed. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow.

Binding and/or activating compounds can also be screened by using chimeric receptor proteins in which the amino terminal extracellular domain, or parts thereof, the entire transmembrane domain or subregions, such as any of the seven transmembrane segments or any of the intracellular or extracellular loops and the carboxy terminal intracellular domain, or parts thereof, can be replaced by heterologous domains or subregions. For example, a G-protein-binding region can be used that interacts with a different G-protein then that which is recognized by the native receptor. Accordingly, a different set of signal transduction components is available as an end-point assay for activation. Alternatively, the entire transmembrane portion or subregions (such as transmembrane segments or intracellular or extracellular loops) can be replaced with the entire transmembrane portion or subregions specific to a host cell that is different from the host cell from which the amino terminal extracellular domain and/or the G-protein-binding region are derived. This allows for assays to be performed in other than the specific host cell from which the receptor is derived. Alternatively, the amino terminal extracellular domain (and/or other ligand-binding regions) could be replaced by a domain (and/or other binding region) binding a different ligand, thus, providing an assay for test compounds that interact with the heterologous amino terminal extracellular domain (or region) but still cause signal transduction. Finally, activation can be detected by a reporter gene containing an easily detectable coding region operably linked to a transcriptional regulatory sequence that is part of the native signal transduction pathway.

The proteins of the present invention are also useful in competition binding assays in methods designed to discover compounds that interact with the receptor. Thus, a compound is exposed to a receptor polypeptide under conditions that allow the compound to bind or to otherwise interact with the polypeptide (Hodgson, Bio/technology, 1992, Sept 10(9);973-80). Soluble receptor polypeptide is also added to the mixture. If the test compound interacts with the soluble receptor polypeptide, it decreases the amount of complex formed or activity from the

receptor target. This type of assay is particularly useful in cases in which compounds are sought that interact with specific regions of the receptor. Thus, the soluble polypeptide that competes with the target receptor region is designed to contain peptide sequences corresponding to the region of interest.

To perform cell free drug screening assays, it is sometimes desirable to immobilize either the receptor protein, or fragment, or its target molecule to facilitate separation of complexes from uncomplexed forms of one or both of the proteins, as well as to accommodate automation of the assay.

Techniques for immobilizing proteins on matrices can be used in the drug screening assays. In one embodiment, a fusion protein can be provided which adds a domain that allows the protein to be bound to a matrix. For example, glutathione-S-transferase fusion proteins can be adsorbed onto glutathione sepharose beads (Sigma Chemical, St. Louis, MO) or glutathione derivatized microtitre plates, which are then combined with the cell lysates (e.g., ^{35}S-labeled) and the candidate compound, and the mixture incubated under conditions conducive to complex formation (e.g., at physiological conditions for salt and pH). Following incubation, the beads are washed to remove any unbound label, and the matrix immobilized and radiolabel determined directly, or in the supernatant after the complexes are dissociated. Alternatively, the complexes can be dissociated from the matrix, separated by SDS-PAGE, and the level of receptor-binding protein found in the bead fraction quantitated from the gel using standard electrophoretic techniques. For example, either the polypeptide or its target molecule can be immobilized utilizing conjugation of biotin and streptavidin using techniques well known in the art. Alternatively, antibodies reactive with the protein but which do not interfere with binding of the protein to its target molecule can be derivatized to the wells of the plate, and the protein trapped in the wells by antibody conjugation. Preparations of a receptor-binding protein and a candidate compound are incubated in the receptor protein-presenting wells and the amount of complex trapped in the well can be quantitated. Methods for detecting such complexes, in addition to those described above for the GST-immobilized complexes, include immunodetection of complexes using antibodies reactive with the receptor protein target molecule, or which are reactive with receptor protein and compete with the target molecule, as well as enzyme-linked assays which rely on detecting an enzymatic activity associated with the target molecule.

Agents that modulate one of the GPCRs of the present invention can be identified using one or more of the above assays, alone or in combination. It is generally preferable to use a cell-based or cell free system first and then confirm activity in an animal or other model system. Such model systems are well known in the art and can readily be employed in this context.

Modulators of receptor protein activity identified according to these drug screening assays can be used to treat a subject with a disorder mediated by the receptor pathway, by treating cells or tissues that express the GPCR. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. These methods of treatment include the steps of administering a modulator of the GPCR's activity in a pharmaceutical composition to a subject in need of such treatment, the modulator being identified as described herein.

In yet another aspect of the invention, the GPCR proteins can be used as "bait proteins" in a two-hybrid assay or three-hybrid assay (see, e.g., U.S. Patent No. 5,283,317; Zervos et al. (1993) *Cell* 72:223-232; Madura et al. (1993) *J. Biol. Chem.* 268:12046-12054; Bartel et al. (1993) *Biotechniques* 14:920-924; Iwabuchi et al. (1993) *Oncogene* 8:1693-1696; and Brent WO94/10300), to identify other proteins, which bind to or interact with the GPCR and are involved in GPCR activity. Such GPCR-binding proteins are also likely to be involved in the propagation of signals by the GPCR proteins or GPCR targets as, for example, downstream elements of a GPCR-mediated signaling pathway. Alternatively, such GPCR-binding proteins are likely to be GPCR inhibitors.

The two-hybrid system is based on the modular nature of most transcription factors, which consist of separable DNA-binding and activation domains. Briefly, the assay utilizes two different DNA constructs. In one construct, the gene that codes for a GPCR protein is fused to a gene encoding the DNA binding domain of a known transcription factor (e.g., GAL-4). In the other construct, a DNA sequence, from a library of DNA sequences, that encodes an unidentified protein ("prey" or "sample") is fused to a gene that codes for the activation domain of the known transcription factor. If the "bait" and the "prey" proteins are able to interact, *in vivo,* forming a GPCR-dependent complex, the DNA-binding and activation domains of the transcription factor are brought into close proximity. This proximity allows transcription of a reporter gene (e.g., LacZ) which is operably linked to a transcriptional regulatory site responsive to the transcription factor. Expression of the reporter gene can be detected and cell colonies containing the functional transcription factor can be isolated and used to obtain the cloned gene which encodes the protein which interacts with the GPCR protein.

This invention further pertains to novel agents identified by the above-described screening assays. Accordingly, it is within the scope of this invention to further use an agent identified as described herein in an appropriate animal model. For example, an agent identified as described herein (e.g., a GPCR modulating agent, an antisense GPCR nucleic

acid molecule, a GPCR-specific antibody, or a GPCR-binding partner) can be used in an animal or other model to determine the efficacy, toxicity, or side effects of treatment with such an agent. Alternatively, an agent identified as described herein can be used in an animal or other model to determine the mechanism of action of such an agent. Furthermore, this invention pertains to uses of novel agents identified by the above-described screening assays for treatments as described herein.

The GPCR proteins of the present invention are also useful to provide a target for diagnosing a disease or predisposition to disease mediated by the peptide. Accordingly, the invention provides methods for detecting the presence, or levels of, the protein (or encoding mRNA) in a cell, tissue, or organism. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. The method involves contacting a biological sample with a compound capable of interacting with the receptor protein such that the interaction can be detected. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

One agent for detecting a protein in a sample is an antibody capable of selectively binding to protein. A biological sample includes tissues, cells and biological fluids isolated from a subject, as well as tissues, cells and fluids present within a subject.

The peptides of the present invention also provide targets for diagnosing active protein activity, disease, or predisposition to disease, in a patient having a variant peptide, particularly activities and conditions that are known for other members of the family of proteins to which the present one belongs. Thus, the peptide can be isolated from a biological sample and assayed for the presence of a genetic mutation that results in aberrant peptide. This includes amino acid substitution, deletion, insertion, rearrangement, (as the result of aberrant splicing events), and inappropriate post-translational modification. Analytic methods include altered electrophoretic mobility, altered tryptic peptide digest, altered receptor activity in cell-based or cell-free assay, alteration in ligand or antibody-binding pattern, altered isoelectric point, direct amino acid sequencing, and any other of the known assay techniques useful for detecting mutations in a protein. Such an assay can be provided in a single detection format or a multi-detection format such as an antibody chip array.

In vitro techniques for detection of peptide include enzyme linked immunosorbent assays (ELISAs), Western blots, immunoprecipitations and immunofluorescence using a detection reagent, such as an antibody or protein binding agent. Alternatively, the peptide can be detected *in vivo* in a subject by introducing into the subject a labeled anti-peptide antibody or other types of detection agent. For example, the antibody can be labeled with a radioactive

marker whose presence and location in a subject can be detected by standard imaging techniques. Particularly useful are methods that detect the allelic variant of a peptide expressed in a subject and methods which detect fragments of a peptide in a sample.

The peptides are also useful in pharmacogenomic analysis. Pharmacogenomics deal with clinically significant hereditary variations in the response to drugs due to altered drug disposition and abnormal action in affected persons. See, e.g., Eichelbaum, M. (*Clin. Exp. Pharmacol. Physiol.* 23(10-11):983-985 (1996)), and Linder, M.W. (*Clin. Chem.* 43(2):254-266 (1997)). The clinical outcomes of these variations result in severe toxicity of therapeutic drugs in certain individuals or therapeutic failure of drugs in certain individuals as a result of individual variation in metabolism. Thus, the genotype of the individual can determine the way a therapeutic compound acts on the body or the way the body metabolizes the compound. Further, the activity of drug metabolizing enzymes effects both the intensity and duration of drug action. Thus, the pharmacogenomics of the individual permit the selection of effective compounds and effective dosages of such compounds for prophylactic or therapeutic treatment based on the individual's genotype. The discovery of genetic polymorphisms in some drug metabolizing enzymes has explained why some patients do not obtain the expected drug effects, show an exaggerated drug effect, or experience serious toxicity from standard drug dosages. Polymorphisms can be expressed in the phenotype of the extensive metabolizer and the phenotype of the poor metabolizer. Accordingly, genetic polymorphism may lead to allelic protein variants of the receptor protein in which one or more of the receptor functions in one population is different from those in another population. The peptides thus allow a target to ascertain a genetic predisposition that can affect treatment modality. Thus, in a ligand-based treatment, polymorphism may give rise to amino terminal extracellular domains and/or other ligand-binding regions that are more or less active in ligand binding, and receptor activation. Accordingly, ligand dosage would necessarily be modified to maximize the therapeutic effect within a given population containing a polymorphism. As an alternative to genotyping, specific polymorphic peptides could be identified.

The peptides are also useful for treating a disorder characterized by an absence of, inappropriate, or unwanted expression of the protein. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Accordingly, methods for treatment include the use of the GPCR protein or fragments.

Antibodies

The invention also provides antibodies that selectively bind to one of the peptides of the present invention, a protein comprising such a peptide, as well as variants and fragments thereof. As used herein, an antibody selectively binds a target peptide when it binds the target peptide and does not significantly bind to unrelated proteins. An antibody is still considered to selectively bind a peptide even if it also binds to other proteins that are not substantially homologous with the target peptide so long as such proteins share homology with a fragment or domain of the peptide target of the antibody. In this case, it would be understood that antibody binding to the peptide is still selective despite some degree of cross-reactivity.

As used herein, an antibody is defined in terms consistent with that recognized within the art: they are multi-subunit proteins produced by a mammalian organism in response to an antigen challenge. The antibodies of the present invention include polyclonal antibodies and monoclonal antibodies, as well as fragments of such antibodies, including, but not limited to, Fab or $F(ab')_2$, and Fv fragments.

Many methods are known for generating and/or identifying antibodies to a given target peptide. Several such methods are described by Harlow, Antibodies, Cold Spring Harbor Press, (1989).

In general, to generate antibodies, an isolated peptide is used as an immunogen and is administered to a mammalian organism, such as a rat, rabbit or mouse. The full-length protein, an antigenic peptide fragment or a fusion protein can be used. Particularly important fragments are those covering functional domains, such as the domains identified in Figure 2, and domain of sequence homology or divergence amongst the family, such as those that can readily be identified using protein alignment methods and as presented in the Figures.

Antibodies are preferably prepared from regions or discrete fragments of the GPCR proteins. Antibodies can be prepared from any region of the peptide as described herein. However, preferred regions will include those involved in function/activity and/or receptor/binding partner interaction. Figure 2 can be used to identify particularly important regions while sequence alignment can be used to identify conserved and unique sequence fragments.

An antigenic fragment will typically comprise at least 8 contiguous amino acid residues. The antigenic peptide can comprise, however, at least 10, 12, 14, 16 or more amino acid residues. Such fragments can be selected on a physical property, such as fragments correspond

to regions that are located on the surface of the protein, e.g., hydrophilic regions or can be selected based on sequence uniqueness (see Figure 2).

Detection on an antibody of the present invention can be facilitated by coupling (i.e., physically linking) the antibody to a detectable substance. Examples of detectable substances include various enzymes, prosthetic groups, fluorescent materials, luminescent materials, bioluminescent materials, and radioactive materials. Examples of suitable enzymes include horseradish peroxidase, alkaline phosphatase, β-galactosidase, or acetylcholinesterase; examples of suitable prosthetic group complexes include streptavidin/biotin and avidin/biotin; examples of suitable fluorescent materials include umbelliferone, fluorescein, fluorescein isothiocyanate, rhodamine, dichlorotriazinylamine fluorescein, dansyl chloride or phycoerythrin; an example of a luminescent material includes luminol; examples of bioluminescent materials include luciferase, luciferin, and aequorin, and examples of suitable radioactive material include ^{125}I, ^{131}I, ^{35}S or ^{3}H.

Antibody Uses

The antibodies can be used to isolate one of the proteins of the present invention by standard techniques, such as affinity chromatography or immunoprecipitation. The antibodies can facilitate the purification of the natural protein from cells and recombinantly produced protein expressed in host cells. In addition, such antibodies are useful to detect the presence of one of the proteins of the present invention in cells or tissues to determine the pattern of expression of the protein among various tissues in an organism and over the course of normal development. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow. Further, such antibodies can be used to detect protein *in situ*, *in vitro*, or in a cell lysate or supernatant in order to evaluate the abundance and pattern of expression. Also, such antibodies can be used to assess abnormal tissue distribution or abnormal expression during development or progression of a biological condition. Antibody detection of circulating fragments of the full length protein can be used to identify turnover.

Further, the antibodies can be used to assess expression in disease states such as in active stages of the disease or in an individual with a predisposition toward disease related to the protein's function. When a disorder is caused by an inappropriate tissue distribution,

developmental expression, level of expression of the protein, or expressed/processed form, the antibody can be prepared against the normal protein. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. If a disorder is characterized by a specific mutation in the protein, antibodies specific for this mutant protein can be used to assay for the presence of the specific mutant protein.

The antibodies can also be used to assess normal and aberrant subcellular localization of cells in the various tissues in an organism. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. The diagnostic uses can be applied, not only in genetic testing, but also in monitoring a treatment modality. Accordingly, where treatment is ultimately aimed at correcting expression level or the presence of aberrant sequence and aberrant tissue distribution or developmental expression, antibodies directed against the protein or relevant fragments can be used to monitor therapeutic efficacy.

Additionally, antibodies are useful in pharmacogenomic analysis. Thus, antibodies prepared against polymorphic proteins can be used to identify individuals that require modified treatment modalities. The antibodies are also useful as diagnostic tools as an immunological marker for aberrant protein analyzed by electrophoretic mobility, isoelectric point, tryptic peptide digest, and other physical assays known to those in the art.

The antibodies are also useful for tissue typing. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Thus, where a specific protein has been correlated with expression in a specific tissue, antibodies that are specific for this protein can be used to identify a tissue type.

The antibodies are also useful for inhibiting protein function, for example, blocking the binding of the GPCR peptide to a binding partner such as a ligand. These uses can also be applied in a therapeutic context in which treatment involves inhibiting the protein's function. An antibody can be used, for example, to block binding, thus modulating (agonizing or antagonizing) the peptides activity. Antibodies can be prepared against specific fragments containing sites required for function or against intact protein that is associated with a cell or cell membrane. See Figure 2 for structural information relating to the proteins of the present invention.

The invention also encompasses kits for using antibodies to detect the presence of a protein in a biological sample. The kit can comprise antibodies such as a labeled or labelable antibody and a compound or agent for detecting protein in a biological sample; means for determining the amount of protein in the sample; means for comparing the amount of protein in

the sample with a standard; and instructions for use. Such a kit can be supplied to detect a single protein or epitope or can be configured to detect one of a multitude of epitopes, such as in an antibody detection array. Arrays are described in detail below for nucleic acid arrays and similar methods have been developed for antibody arrays.

Nucleic Acid Molecules

The present invention further provides isolated nucleic acid molecules that encode a GPCR peptide or protein of the present invention (cDNA, transcript and genomic sequence). Such nucleic acid molecules will consist of, consist essentially of, or comprise a nucleotide sequence that encodes one of the GPCR peptides of the present invention, an allelic variant thereof, or an ortholog or paralog thereof.

As used herein, an "isolated" nucleic acid molecule is one that is separated from other nucleic acid present in the natural source of the nucleic acid. Preferably, an "isolated" nucleic acid is free of sequences which naturally flank the nucleic acid (i.e., sequences located at the 5' and 3' ends of the nucleic acid) in the genomic DNA of the organism from which the nucleic acid is derived. However, there can be some flanking nucleotide sequences, for example up to about 5KB, 4KB, 3KB, 2KB, or 1KB or less, particularly contiguous peptide encoding sequences and peptide encoding sequences within the same gene but separated by introns in the genomic sequence. The important point is that the nucleic acid is isolated from remote and unimportant flanking sequences such that it can be subjected to the specific manipulations described herein such as recombinant expression, preparation of probes and primers, and other uses specific to the nucleic acid sequences.

Moreover, an "isolated" nucleic acid molecule, such as a transcript/cDNA molecule, can be substantially free of other cellular material, or culture medium when produced by recombinant techniques, or chemical precursors or other chemicals when chemically synthesized. However, the nucleic acid molecule can be fused to other coding or regulatory sequences and still be considered isolated.

For example, recombinant DNA molecules contained in a vector are considered isolated. Further examples of isolated DNA molecules include recombinant DNA molecules maintained in heterologous host cells or purified (partially or substantially) DNA molecules in solution. Isolated RNA molecules include *in vivo* or *in vitro* RNA transcripts of the isolated DNA molecules of the present invention. Isolated nucleic acid molecules according to the present invention further include such molecules produced synthetically.

Accordingly, the present invention provides nucleic acid molecules that consist of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, cDNA sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists of a nucleotide sequence when the nucleotide sequence is the complete nucleotide sequence of the nucleic acid molecule.

The present invention further provides nucleic acid molecules that consist essentially of the nucleotide sequence shown in Figure 1 or 3 (SEQ ID NO:1, cDNA sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule consists essentially of a nucleotide sequence when such a nucleotide sequence is present with only a few additional nucleic acid residues in the final nucleic acid molecule.

The present invention further provides nucleic acid molecules that comprise the nucleotide sequences shown in Figure 1 or 3 (SEQ ID NO:1, cDNA sequence and SEQ ID NO:3, genomic sequence), or any nucleic acid molecule that encodes the protein provided in Figure 2, SEQ ID NO:2. A nucleic acid molecule comprises a nucleotide sequence when the nucleotide sequence is at least part of the final nucleotide sequence of the nucleic acid molecule. In such a fashion, the nucleic acid molecule can be only the nucleotide sequence or have additional nucleic acid residues, such as nucleic acid residues that are naturally associated with it or heterologous nucleotide sequences. Such a nucleic acid molecule can have a few additional nucleotides or can comprises several hundred or more additional nucleotides. A brief description of how various types of these nucleic acid molecules can be readily made/isolated is provided below.

In Figures 1 and 3, both coding and non-coding sequences are provided. Because of the source of the present invention, human genomic sequences (Figure 3) and cDNA/transcript sequences (Figure 1), the nucleic acid molecules in the Figures will contain genomic intronic sequences, 5' and 3' non-coding sequences, gene regulatory regions and non-coding intergenic sequences. In general such sequence features are either noted in Figures 1 and 3 or can readily be identified using computational tools known in the art. As discussed below, some of the non-coding regions, particularly gene regulatory elements such as promoters, are useful for a variety of purposes, e.g. control of heterologous gene expression, target for identifying gene activity modulating compounds, and are particularly claimed as fragments of the genomic sequence provided herein.

The isolated nucleic acid molecules can encode the mature protein plus additional amino or carboxyl-terminal amino acids, or amino acids interior to the mature peptide (when the

mature form has more than one peptide chain, for instance). Such sequences may play a role in processing of a protein from precursor to a mature form, facilitate protein trafficking, prolong or shorten protein half-life or facilitate manipulation of a protein for assay or production, among other things. As generally is the case *in situ*, the additional amino acids may be processed away from the mature protein by cellular enzymes.

As mentioned above, the isolated nucleic acid molecules include, but are not limited to, the sequence encoding the GPCR peptide alone, the sequence encoding the mature peptide and additional coding sequences, such as a leader or secretory sequence (e.g., a pre-pro or pro-protein sequence), the sequence encoding the mature peptide, with or without the additional coding sequences, plus additional non-coding sequences, for example introns and non-coding 5' and 3' sequences such as transcribed but non-translated sequences that play a role in transcription, mRNA processing (including splicing and polyadenylation signals), ribosome binding and stability of mRNA. In addition, the nucleic acid molecule may be fused to a marker sequence encoding, for example, a peptide that facilitates purification.

Isolated nucleic acid molecules can be in the form of RNA, such as mRNA, or in the form DNA, including cDNA and genomic DNA obtained by cloning or produced by chemical synthetic techniques or by a combination thereof. The nucleic acid, especially DNA, can be double-stranded or single-stranded. Single-stranded nucleic acid can be the coding strand (sense strand) or the non-coding strand (anti-sense strand).

The invention further provides nucleic acid molecules that encode fragments of the peptides of the present invention as well as nucleic acid molecules that encode obvious variants of the GPCR proteins of the present invention that are described above. Such nucleic acid molecules may be naturally occurring, such as allelic variants (same locus), paralogs (different locus), and orthologs (different organism), or may be constructed by recombinant DNA methods or by chemical synthesis. Such non-naturally occurring variants may be made by mutagenesis techniques, including those applied to nucleic acid molecules, cells, or organisms. Accordingly, as discussed above, the variants can contain nucleotide substitutions, deletions, inversions and insertions. Variation can occur in either or both the coding and non-coding regions. The variations can produce both conservative and non-conservative amino acid substitutions.

The present invention further provides non-coding fragments of the nucleic acid molecules provided in Figures 1 and 3. Preferred non-coding fragments include, but are not limited to, promoter sequences, enhancer sequences, gene modulating sequences and gene termination sequences. Such fragments are useful in controlling heterologous gene expression

and in developing screens to identify gene-modulating agents. A promoter can readily be identified as being 5' to the ATG start site in the genomic sequence provided in Figure 3.

A fragment comprises a contiguous nucleotide sequence greater than 12 or more nucleotides. Further, a fragment could at least 30, 40, 50, 100, 250 or 500 nucleotides in length. The length of the fragment will be based on its intended use. For example, the fragment can encode epitope bearing regions of the peptide, or can be useful as DNA probes and primers. Such fragments can be isolated using the known nucleotide sequence to synthesize an oligonucleotide probe. A labeled probe can then be used to screen a cDNA library, genomic DNA library, or mRNA to isolate nucleic acid corresponding to the coding region. Further, primers can be used in PCR reactions to clone specific regions of gene.

A probe/primer typically comprises substantially a purified oligonucleotide or oligonucleotide pair. The oligonucleotide typically comprises a region of nucleotide sequence that hybridizes under stringent conditions to at least about 12, 20, 25, 40, 50 or more consecutive nucleotides.

Orthologs, homologs, and allelic variants can be identified using methods well known in the art. As described in the Peptide Section, these variants comprise a nucleotide sequence encoding a peptide that is typically 60-70%, 70-80%, 80-90%, and more typically at least about 90-95% or more homologous to the nucleotide sequence shown in the Figure sheets or a fragment of this sequence. Such nucleic acid molecules can readily be identified as being able to hybridize under moderate to stringent conditions, to the nucleotide sequence shown in the Figure sheets or a fragment of the sequence. Allelic variants can readily be determined by genetic locus of the encoding gene. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

Figure 3 provides information on SNPs that have been found in a gene encoding the GPCR proteins of the present invention. The following SNPs were found: T406C, T852C, G897A, C1433T, T5845C, and G7028A.

As used herein, the term "hybridizes under stringent conditions" is intended to describe conditions for hybridization and washing under which nucleotide sequences encoding a peptide at least 60-70% homologous to each other typically remain hybridized to each other. The conditions can be such that sequences at least about 60%, at least about 70%, or at least about 80% or more homologous to each other typically remain hybridized to each other. Such stringent conditions are known to those skilled in the art and can be found in *Current Protocols in Molecular Biology*, John Wiley & Sons, N.Y. (1989), 6.3.1-6.3.6. One example of stringent hybridization conditions are hybridization in 6X sodium chloride/sodium citrate (SSC) at about

45C, followed by one or more washes in 0.2 X SSC, 0.1% SDS at 50-65C. Examples of moderate to low stringency hybridization conditions are well known in the art.

Nucleic Acid Molecule Uses

The nucleic acid molecules of the present invention are useful for probes, primers, chemical intermediates, and in biological assays. The nucleic acid molecules are useful as a hybridization probe for messenger RNA, transcript/cDNA and genomic DNA to isolate full-length cDNA and genomic clones encoding the peptide described in Figure 2 and to isolate cDNA and genomic clones that correspond to variants (alleles, orthologs, etc.) producing the same or related peptides shown in Figure 2. As illustrated in Figure 3, known SNP variations include T406C, T852C, G897A, C1433T, T5845C, and G7028A.

The probe can correspond to any sequence along the entire length of the nucleic acid molecules provided in the Figures. Accordingly, it could be derived from 5' noncoding regions, the coding region, and 3' noncoding regions. However, as discussed, fragments are not to be construed as encompassing fragments disclosed prior to the present invention.

The nucleic acid molecules are also useful as primers for PCR to amplify any given region of a nucleic acid molecule and are useful to synthesize antisense molecules of desired length and sequence.

The nucleic acid molecules are also useful for constructing recombinant vectors. Such vectors include expression vectors that express a portion of, or all of, the peptide sequences. Vectors also include insertion vectors, used to integrate into another nucleic acid molecule sequence, such as into the cellular genome, to alter *in situ* expression of a gene and/or gene product. For example, an endogenous coding sequence can be replaced via homologous recombination with all or part of the coding region containing one or more specifically introduced mutations.

The nucleic acid molecules are also useful for expressing antigenic portions of the proteins.

The nucleic acid molecules are also useful as probes for determining the chromosomal positions of the nucleic acid molecules by means of *in situ* hybridization methods. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1.

The nucleic acid molecules are also useful in making vectors containing the gene regulatory regions of the nucleic acid molecules of the present invention.

The nucleic acid molecules are also useful for designing ribozymes corresponding to all, or a part, of the mRNA produced from the nucleic acid molecules described herein.

The nucleic acid molecules are also useful for making vectors that express part, or all, of the peptides.

The nucleic acid molecules are also useful for constructing host cells expressing a part, or all, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful for constructing transgenic animals expressing all, or a part, of the nucleic acid molecules and peptides.

The nucleic acid molecules are also useful as hybridization probes for determining the presence, level, form and distribution of nucleic acid expression. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow. Accordingly, the probes can be used to detect the presence of, or to determine levels of, a specific nucleic acid molecule in cells, tissues, and in organisms. The nucleic acid whose level is determined can be DNA or RNA. Accordingly, probes corresponding to the peptides described herein can be used to assess expression and/or gene copy number in a given cell, tissue, or organism. These uses are relevant for diagnosis of disorders involving an increase or decrease in GPCR protein expression relative to normal results.

In vitro techniques for detection of mRNA include Northern hybridizations and *in situ* hybridizations. *In vitro* techniques for detecting DNA includes Southern hybridizations and *in situ* hybridization.

Probes can be used as a part of a diagnostic test kit for identifying cells or tissues that express a GPCR protein, such as by measuring a level of a receptor-encoding nucleic acid in a sample of cells from a subject e.g., mRNA or genomic DNA, or determining if a receptor gene has been mutated. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow.

Nucleic acid expression assays are useful for drug screening to identify compounds that modulate GPCR nucleic acid expression.

The invention thus provides a method for identifying a compound that can be used to treat a disorder associated with nucleic acid expression of the GPCR gene, particularly biological and pathological processes that are mediated by the GPCR in cells and tissues that express it. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. The method typically includes assaying the ability of the compound to modulate the expression of the GPCR nucleic acid and thus identifying a compound that can be used to treat a disorder characterized by undesired GPCR nucleic acid expression. The assays can be performed in cell-based and cell-free systems. Cell-based assays include cells naturally expressing the GPCR nucleic acid or recombinant cells genetically engineered to express specific nucleic acid sequences.

The assay for GPCR nucleic acid expression can involve direct assay of nucleic acid levels, such as mRNA levels, or on collateral compounds involved in the signal pathway. Further, the expression of genes that are up- or down-regulated in response to the GPCR protein signal pathway can also be assayed. In this embodiment the regulatory regions of these genes can be operably linked to a reporter gene such as luciferase.

Thus, modulators of GPCR gene expression can be identified in a method wherein a cell is contacted with a candidate compound and the expression of mRNA determined. The level of expression of GPCR mRNA in the presence of the candidate compound is compared to the level of expression of GPCR mRNA in the absence of the candidate compound. The candidate compound can then be identified as a modulator of nucleic acid expression based on this comparison and be used, for example to treat a disorder characterized by aberrant nucleic acid expression. When expression of mRNA is statistically significantly greater in the presence of the candidate compound than in its absence, the candidate compound is identified as a stimulator of nucleic acid expression. When nucleic acid expression is statistically significantly less in the presence of the candidate compound than in its absence, the candidate compound is identified as an inhibitor of nucleic acid expression.

The invention further provides methods of treatment, with the nucleic acid as a target, using a compound identified through drug screening as a gene modulator to modulate GPCR nucleic acid expression, particularly to modulate activities within a cell or tissue that expresses the proteins. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in

humans in Hela cells and bone marrow. Modulation includes both up-regulation (i.e. activation or agonization) or down-regulation (suppression or antagonization) or nucleic acid expression.

Alternatively, a modulator for GPCR nucleic acid expression can be a small molecule or drug identified using the screening assays described herein as long as the drug or small molecule inhibits the GPCR nucleic acid expression in the cells and tissues that express the protein. Experimental data as provided in Figure 1 indicates expression in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells.

The nucleic acid molecules are also useful for monitoring the effectiveness of modulating compounds on the expression or activity of the GPCR gene in clinical trials or in a treatment regimen. Thus, the gene expression pattern can serve as a barometer for the continuing effectiveness of treatment with the compound, particularly with compounds to which a patient can develop resistance. The gene expression pattern can also serve as a marker indicative of a physiological response of the affected cells to the compound. Accordingly, such monitoring would allow either increased administration of the compound or the administration of alternative compounds to which the patient has not become resistant. Similarly, if the level of nucleic acid expression falls below a desirable level, administration of the compound could be commensurately decreased.

The nucleic acid molecules are also useful in diagnostic assays for qualitative changes in GPCR nucleic acid, and particularly in qualitative changes that lead to pathology. The nucleic acid molecules can be used to detect mutations in GPCR genes and gene expression products such as mRNA. The nucleic acid molecules can be used as hybridization probes to detect naturally-occurring genetic mutations in the GPCR gene and thereby to determine whether a subject with the mutation is at risk for a disorder caused by the mutation. Mutations include deletion, addition, or substitution of one or more nucleotides in the gene, chromosomal rearrangement, such as inversion or transposition, modification of genomic DNA, such as aberrant methylation patterns or changes in gene copy number, such as amplification. Detection of a mutated form of the GPCR gene associated with a dysfunction provides a diagnostic tool for an active disease or susceptibility to disease when the disease results from overexpression, underexpression, or altered expression of a GPCR protein.

Individuals carrying mutations in the GPCR gene can be detected at the nucleic acid level by a variety of techniques. Figure 3 provides information on SNPs that have been found in a gene encoding the GPCR proteins of the present invention. The following SNPs were found: T406C, T852C, G897A, C1433T, T5845C, and G7028A. As indicated by the data presented in Figure 3, the map position was determined to be on chromosome 1. Genomic DNA can be

analyzed directly or can be amplified by using PCR prior to analysis. RNA or cDNA can be used in the same way. In some uses, detection of the mutation involves the use of a probe/primer in a polymerase chain reaction (PCR) (see, e.g. U.S. Patent Nos. 4,683,195 and 4,683,202), such as anchor PCR or RACE PCR, or, alternatively, in a ligation chain reaction (LCR) (see, e.g., Landegran *et al.*, *Science 241*:1077-1080 (1988); and Nakazawa *et al.*, *PNAS 91*:360-364 (1994)), the latter of which can be particularly useful for detecting point mutations in the gene (see Abravaya *et al.*, *Nucleic Acids Res. 23*:675-682 (1995)). This method can include the steps of collecting a sample of cells from a patient, isolating nucleic acid (e.g., genomic, mRNA or both) from the cells of the sample, contacting the nucleic acid sample with one or more primers which specifically hybridize to a gene under conditions such that hybridization and amplification of the gene (if present) occurs, and detecting the presence or absence of an amplification product, or detecting the size of the amplification product and comparing the length to a control sample. Deletions and insertions can be detected by a change in size of the amplified product compared to the normal genotype. Point mutations can be identified by hybridizing amplified DNA to normal RNA or antisense DNA sequences.

Alternatively, mutations in a GPCR gene can be directly identified, for example, by alterations in restriction enzyme digestion patterns determined by gel electrophoresis.

Further, sequence-specific ribozymes (U.S. Patent No. 5,498,531) can be used to score for the presence of specific mutations by development or loss of a ribozyme cleavage site. Perfectly matched sequences can be distinguished from mismatched sequences by nuclease cleavage digestion assays or by differences in melting temperature.

Sequence changes at specific locations can also be assessed by nuclease protection assays such as RNase and S1 protection or the chemical cleavage method. Furthermore, sequence differences between a mutant GPCR gene and a wild-type gene can be determined by direct DNA sequencing. A variety of automated sequencing procedures can be utilized when performing the diagnostic assays (Naeve, C.W., (1995) *Biotechniques 19*:448), including sequencing by mass spectrometry (see, e.g., PCT International Publication No. WO 94/16101; Cohen *et al.*, *Adv. Chromatogr. 36*:127-162 (1996); and Griffin *et al.*, *Appl. Biochem. Biotechnol. 38*:147-159 (1993)).

Other methods for detecting mutations in the gene include methods in which protection from cleavage agents is used to detect mismatched bases in RNA/RNA or RNA/DNA duplexes (Myers *et al.*, *Science 230*:1242 (1985)); Cotton *et al.*, *PNAS 85*:4397 (1988); Saleeba *et al.*, *Meth. Enzymol. 217*:286-295 (1992)), electrophoretic mobility of mutant and wild type nucleic acid is compared (Orita *et al.*, *PNAS 86*:2766 (1989); Cotton *et al.*, *Mutat. Res. 285*:125-144

(1993); and Hayashi *et al.*, *Genet. Anal. Tech. Appl.* *9*:73-79 (1992)), and movement of mutant or wild-type fragments in polyacrylamide gels containing a gradient of denaturant is assayed using denaturing gradient gel electrophoresis (Myers *et al.*, *Nature* *313*:495 (1985)). Examples of other techniques for detecting point mutations include selective oligonucleotide hybridization, selective amplification, and selective primer extension.

The nucleic acid molecules are also useful for testing an individual for a genotype that while not necessarily causing the disease, nevertheless affects the treatment modality. Thus, the nucleic acid molecules can be used to study the relationship between an individual's genotype and the individual's response to a compound used for treatment (pharmacogenomic relationship). Accordingly, the nucleic acid molecules described herein can be used to assess the mutation content of the GPCR gene in an individual in order to select an appropriate compound or dosage regimen for treatment. As illustrated in Figure 3, known SNP variations include T406C, T852C, G897A, C1433T, T5845C, and G7028A.

Thus nucleic acid molecules displaying genetic variations that affect treatment provide a diagnostic target that can be used to tailor treatment in an individual. Accordingly, the production of recombinant cells and animals containing these polymorphisms allow effective clinical design of treatment compounds and dosage regimens.

The nucleic acid molecules are thus useful as antisense constructs to control GPCR gene expression in cells, tissues, and organisms. A DNA antisense nucleic acid molecule is designed to be complementary to a region of the gene involved in transcription, preventing transcription and hence production of GPCR protein. An antisense RNA or DNA nucleic acid molecule would hybridize to the mRNA and thus block translation of mRNA into GPCR protein.

Alternatively, a class of antisense molecules can be used to inactivate mRNA in order to decrease expression of GPCR nucleic acid. Accordingly, these molecules can treat a disorder characterized by abnormal or undesired GPCR nucleic acid expression. This technique involves cleavage by means of ribozymes containing nucleotide sequences complementary to one or more regions in the mRNA that attenuate the ability of the mRNA to be translated. Possible regions include coding regions and particularly coding regions corresponding to the catalytic and other functional activities of the GPCR protein, such as ligand binding.

The nucleic acid molecules also provide vectors for gene therapy in patients containing cells that are aberrant in GPCR gene expression. Thus, recombinant cells, which include the patient's cells that have been engineered *ex vivo* and returned to the patient, are introduced into an individual where the cells produce the desired GPCR protein to treat the individual.

The invention also encompasses kits for detecting the presence of a GPCR nucleic acid in a biological sample. Experimental data as provided in Figure 1 indicates that GPCR proteins of the present invention are expressed in Hela cells, bone marrow, and a pooled sample of fetal lung, testis, and B-cells. Specifically, a virtual northern blot shows expression in a pooled fetal lung/testis/B-cell sample. In addition, PCR-based tissue screening panels indicate expression in humans in Hela cells and bone marrow. For example, the kit can comprise reagents such as a labeled or labelable nucleic acid or agent capable of detecting GPCR nucleic acid in a biological sample; means for determining the amount of GPCR nucleic acid in the sample; and means for comparing the amount of GPCR nucleic acid in the sample with a standard. The compound or agent can be packaged in a suitable container. The kit can further comprise instructions for using the kit to detect GPCR protein mRNA or DNA.

Nucleic Acid Arrays

The present invention further provides nucleic acid detection kits, such as arrays or microarrays of nucleic acid molecules that are based on the sequence information provided in Figures 1 and 3 (SEQ ID NOS:1 and 3).

As used herein "Arrays" or "Microarrays" refers to an array of distinct polynucleotides or oligonucleotides synthesized on a substrate, such as paper, nylon or other type of membrane, filter, chip, glass slide, or any other suitable solid support. In one embodiment, the microarray is prepared and used according to the methods described in US Patent 5,837,832, Chee et al., PCT application W095/11995 (Chee et al.), Lockhart, D. J. et al. (1996; Nat. Biotech. 14: 1675-1680) and Schena, M. et al. (1996; Proc. Natl. Acad. Sci. 93: 10614-10619), all of which are incorporated herein in their entirety by reference. In other embodiments, such arrays are produced by the methods described by Brown et. al., US Patent No. 5,807,522.

The microarray or detection kit is preferably composed of a large number of unique, single-stranded nucleic acid sequences, usually either synthetic antisense oligonucleotides or fragments of cDNAs, fixed to a solid support. The oligonucleotides are preferably about 6-60 nucleotides in length, more preferably 15-30 nucleotides in length, and most preferably about 20-25 nucleotides in length. For a certain type of microarray or detection kit, it may be preferable to use oligonucleotides that are only 7-20 nucleotides in length. The microarray or detection kit may contain oligonucleotides that cover the known 5', or 3', sequence, sequential oligonucleotides which cover the full length sequence; or unique oligonucleotides selected

from particular areas along the length of the sequence. Polynucleotides used in the microarray or detection kit may be oligonucleotides that are specific to a gene or genes of interest.

In order to produce oligonucleotides to a known sequence for a microarray or detection kit, the gene(s) of interest (or an ORF identified from the contigs of the present invention) is typically examined using a computer algorithm which starts at the 5' or at the 3' end of the nucleotide sequence. Typical algorithms will then identify oligomers of defined length that are unique to the gene, have a GC content within a range suitable for hybridization, and lack predicted secondary structure that may interfere with hybridization. In certain situations it may be appropriate to use pairs of oligonucleotides on a microarray or detection kit. The "pairs" will be identical, except for one nucleotide that preferably is located in the center of the sequence. The second oligonucleotide in the pair (mismatched by one) serves as a control. The number of oligonucleotide pairs may range from two to one million. The oligomers are synthesized at designated areas on a substrate using a light-directed chemical process. The substrate may be paper, nylon or other type of membrane, filter, chip, glass slide or any other suitable solid support.

In another aspect, an oligonucleotide may be synthesized on the surface of the substrate by using a chemical coupling procedure and an ink jet application apparatus, as described in PCT application W095/251116 (Baldeschweiler et al.) which is incorporated herein in its entirety by reference. In another aspect, a "gridded" array analogous to a dot (or slot) blot may be used to arrange and link cDNA fragments or oligonucleotides to the surface of a substrate using a vacuum system, thermal, UV, mechanical or chemical bonding procedures. An array, such as those described above, may be produced by hand or by using available devices (slot blot or dot blot apparatus), materials (any suitable solid support), and machines (including robotic instruments), and may contain 8, 24, 96, 384, 1536, 6144 or more oligonucleotides, or any other number between two and one million which lends itself to the efficient use of commercially available instrumentation.

In order to conduct sample analysis using a microarray or detection kit, the RNA or DNA from a biological sample is made into hybridization probes. The mRNA is isolated, and cDNA is produced and used as a template to make antisense RNA (aRNA). The aRNA is amplified in the presence of fluorescent nucleotides, and labeled probes are incubated with the microarray or detection kit so that the probe sequences hybridize to complementary oligonucleotides of the microarray or detection kit. Incubation conditions are adjusted so that hybridization occurs with precise complementary matches or with various degrees of less complementarity. After removal of nonhybridized probes, a scanner is used to determine the

levels and patterns of fluorescence. The scanned images are examined to determine degree of complementarity and the relative abundance of each oligonucleotide sequence on the microarray or detection kit. The biological samples may be obtained from any bodily fluids (such as blood, urine, saliva, phlegm, gastric juices, etc.), cultured cells, biopsies, or other tissue preparations. A detection system may be used to measure the absence, presence, and amount of hybridization for all of the distinct sequences simultaneously. This data may be used for large scale correlation studies on the sequences, expression patterns, mutations, variants, or polymorphisms among samples.

Using such arrays, the present invention provides methods to identify the expression of the GPCR proteins/peptides of the present invention. In detail, such methods comprise incubating a test sample with one or more nucleic acid molecules and assaying for binding of the nucleic acid molecule with components within the test sample. Such assays will typically involve arrays comprising many genes, at least one of which is a gene of the present invention and or alleles of the GPCR gene of the present invention. Figure 3 provides information on SNPs that have been found in a gene encoding the GPCR proteins of the present invention. The following SNPs were found: T406C, T852C, G897A, C1433T, T5845C, and G7028A.

Conditions for incubating a nucleic acid molecule with a test sample vary. Incubation conditions depend on the format employed in the assay, the detection methods employed, and the type and nature of the nucleic acid molecule used in the assay. One skilled in the art will recognize that any one of the commonly available hybridization, amplification or array assay formats can readily be adapted to employ the novel fragments of the Human genome disclosed herein. Examples of such assays can be found in Chard, T, *An Introduction to Radioimmunoassay and Related Techniques,* Elsevier Science Publishers, Amsterdam, The Netherlands (1986); Bullock, G. R. *et al., Techniques in Immunocytochemistry,* Academic Press, Orlando, FL Vol. 1 (1 982), Vol. 2 (1983), Vol. 3 (1985); Tijssen, P., *Practice and Theory of Enzyme Immunoassays: Laboratory Techniques in Biochemistry and Molecular Biology,* Elsevier Science Publishers, Amsterdam, The Netherlands (1985).

The test samples of the present invention include cells, protein or membrane extracts of cells. The test sample used in the above-described method will vary based on the assay format, nature of the detection method and the tissues, cells or extracts used as the sample to be assayed. Methods for preparing nucleic acid extracts or of cells are well known in the art and can be readily be adapted in order to obtain a sample that is compatible with the system utilized.

In another embodiment of the present invention, kits are provided which contain the necessary reagents to carry out the assays of the present invention.

Specifically, the invention provides a compartmentalized kit to receive, in close confinement, one or more containers which comprises: (a) a first container comprising one of the nucleic acid molecules that can bind to a fragment of the Human genome disclosed herein; and (b) one or more other containers comprising one or more of the following: wash reagents, reagents capable of detecting presence of a bound nucleic acid.

In detail, a compartmentalized kit includes any kit in which reagents are contained in separate containers. Such containers include small glass containers, plastic containers, strips of plastic, glass or paper, or arraying material such as silica. Such containers allows one to efficiently transfer reagents from one compartment to another compartment such that the samples and reagents are not cross-contaminated, and the agents or solutions of each container can be added in a quantitative fashion from one compartment to another. Such containers will include a container which will accept the test sample, a container which contains the nucleic acid probe, containers which contain wash reagents (such as phosphate buffered saline, Tris-buffers, etc.), and containers which contain the reagents used to detect the bound probe. One skilled in the art will readily recognize that the previously unidentified GPCR genes of the present invention can be routinely identified using the sequence information disclosed herein can be readily incorporated into one of the established kit formats which are well known in the art, particularly expression arrays.

Vectors/host cells

The invention also provides vectors containing the nucleic acid molecules described herein. The term "vector" refers to a vehicle, preferably a nucleic acid molecule, which can transport the nucleic acid molecules. When the vector is a nucleic acid molecule, the nucleic acid molecules are covalently linked to the vector nucleic acid. With this aspect of the invention, the vector includes a plasmid, single or double stranded phage, a single or double stranded RNA or DNA viral vector, or artificial chromosome, such as a BAC, PAC, YAC, OR MAC.

A vector can be maintained in the host cell as an extrachromosomal element where it replicates and produces additional copies of the nucleic acid molecules. Alternatively, the vector may integrate into the host cell genome and produce additional copies of the nucleic acid molecules when the host cell replicates.

The invention provides vectors for the maintenance (cloning vectors) or vectors for expression (expression vectors) of the nucleic acid molecules. The vectors can function in procaryotic or eukaryotic cells or in both (shuttle vectors).

Expression vectors contain cis-acting regulatory regions that are operably linked in the vector to the nucleic acid molecules such that transcription of the nucleic acid molecules is allowed in a host cell. The nucleic acid molecules can be introduced into the host cell with a separate nucleic acid molecule capable of affecting transcription. Thus, the second nucleic acid molecule may provide a trans-acting factor interacting with the cis-regulatory control region to allow transcription of the nucleic acid molecules from the vector. Alternatively, a trans-acting factor may be supplied by the host cell. Finally, a trans-acting factor can be produced from the vector itself. It is understood, however, that in some embodiments, transcription and/or translation of the nucleic acid molecules can occur in a cell-free system.

The regulatory sequence to which the nucleic acid molecules described herein can be operably linked include promoters for directing mRNA transcription. These include, but are not limited to, the left promoter from bacteriophage λ, the lac, TRP, and TAC promoters from *E. coli*, the early and late promoters from SV40, the CMV immediate early promoter, the adenovirus early and late promoters, and retrovirus long-terminal repeats.

In addition to control regions that promote transcription, expression vectors may also include regions that modulate transcription, such as repressor binding sites and enhancers. Examples include the SV40 enhancer, the cytomegalovirus immediate early enhancer, polyoma enhancer, adenovirus enhancers, and retrovirus LTR enhancers.

In addition to containing sites for transcription initiation and control, expression vectors can also contain sequences necessary for transcription termination and, in the transcribed region a ribosome binding site for translation. Other regulatory control elements for expression include initiation and termination codons as well as polyadenylation signals. The person of ordinary skill in the art would be aware of the numerous regulatory sequences that are useful in expression vectors. Such regulatory sequences are described, for example, in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

A variety of expression vectors can be used to express a nucleic acid molecule. Such vectors include chromosomal, episomal, and virus-derived vectors, for example vectors derived from bacterial plasmids, from bacteriophage, from yeast episomes, from yeast chromosomal elements, including yeast artificial chromosomes, from viruses such as baculoviruses,

papovaviruses such as SV40, Vaccinia viruses, adenoviruses, poxviruses, pseudorabies viruses, and retroviruses. Vectors may also be derived from combinations of these sources such as those derived from plasmid and bacteriophage genetic elements, eg. cosmids and phagemids. Appropriate cloning and expression vectors for prokaryotic and eukaryotic hosts are described in Sambrook *et al.*, *Molecular Cloning: A Laboratory Manual. 2nd. ed.*, Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, (1989).

The regulatory sequence may provide constitutive expression in one or more host cells (i.e. tissue specific) or may provide for inducible expression in one or more cell types such as by temperature, nutrient additive, or exogenous factor such as a hormone or other ligand. A variety of vectors providing for constitutive and inducible expression in prokaryotic and eukaryotic hosts are well known to those of ordinary skill in the art.

The nucleic acid molecules can be inserted into the vector nucleic acid by well-known methodology. Generally, the DNA sequence that will ultimately be expressed is joined to an expression vector by cleaving the DNA sequence and the expression vector with one or more restriction enzymes and then ligating the fragments together. Procedures for restriction enzyme digestion and ligation are well known to those of ordinary skill in the art.

The vector containing the appropriate nucleic acid molecule can be introduced into an appropriate host cell for propagation or expression using well-known techniques. Bacterial cells include, but are not limited to, *E. coli*, *Streptomyces*, and *Salmonella typhimurium*. Eukaryotic cells include, but are not limited to, yeast, insect cells such as *Drosophila*, animal cells such as COS and CHO cells, and plant cells.

As described herein, it may be desirable to express the peptide as a fusion protein. Accordingly, the invention provides fusion vectors that allow for the production of the peptides. Fusion vectors can increase the expression of a recombinant protein, increase the solubility of the recombinant protein, and aid in the purification of the protein by acting for example as a ligand for affinity purification. A proteolytic cleavage site may be introduced at the junction of the fusion moiety so that the desired peptide can ultimately be separated from the fusion moiety. Proteolytic enzymes include, but are not limited to, factor Xa, thrombin, and enterokinase. Typical fusion expression vectors include pGEX (Smith *et al.*, *Gene 67*:31-40 (1988)), pMAL (New England Biolabs, Beverly, MA) and pRIT5 (Pharmacia, Piscataway, NJ) which fuse glutathione S-transferase (GST), maltose E binding protein, or protein A, respectively, to the target recombinant protein. Examples of suitable inducible non-fusion *E. coli* expression vectors include pTrc (Amann *et al.*, *Gene 69*:301-315 (1988)) and pET 11d (Studier *et al.*, *Gene Expression Technology: Methods in Enzymology 185*:60-89 (1990)).

Recombinant protein expression can be maximized in a host bacteria by providing a genetic background wherein the host cell has an impaired capacity to proteolytically cleave the recombinant protein. (Gottesman, S., *Gene Expression Technology: Methods in Enzymology* 185, Academic Press, San Diego, California (1990) 119-128). Alternatively, the sequence of the nucleic acid molecule of interest can be altered to provide preferential codon usage for a specific host cell, for example *E. coli*. (Wada *et al.*, *Nucleic Acids Res. 20*:2111-2118 (1992)).

The nucleic acid molecules can also be expressed by expression vectors that are operative in yeast. Examples of vectors for expression in yeast e.g., *S. cerevisiae* include pYepSec1 (Baldari, *et al.*, *EMBO J. 6*:229-234 (1987)), pMFa (Kurjan *et al.*, *Cell 30*:933-943(1982)), pJRY88 (Schultz *et al.*, *Gene 54*:113-123 (1987)), and pYES2 (Invitrogen Corporation, San Diego, CA).

The nucleic acid molecules can also be expressed in insect cells using, for example, baculovirus expression vectors. Baculovirus vectors available for expression of proteins in cultured insect cells (e.g., Sf 9 cells) include the pAc series (Smith *et al.*, *Mol. Cell Biol. 3*:2156-2165 (1983)) and the pVL series (Lucklow *et al.*, *Virology 170*:31-39 (1989)).

In certain embodiments of the invention, the nucleic acid molecules described herein are expressed in mammalian cells using mammalian expression vectors. Examples of mammalian expression vectors include pCDM8 (Seed, B. *Nature 329*:840(1987)) and pMT2PC (Kaufman *et al.*, *EMBO J. 6*:187-195 (1987)).

The expression vectors listed herein are provided by way of example only of the well-known vectors available to those of ordinary skill in the art that would be useful to express the nucleic acid molecules. The person of ordinary skill in the art would be aware of other vectors suitable for maintenance propagation or expression of the nucleic acid molecules described herein. These are found for example in Sambrook, J., Fritsh, E. F., and Maniatis, T. *Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989.

The invention also encompasses vectors in which the nucleic acid sequences described herein are cloned into the vector in reverse orientation, but operably linked to a regulatory sequence that permits transcription of antisense RNA. Thus, an antisense transcript can be produced to all, or to a portion, of the nucleic acid molecule sequences described herein, including both coding and non-coding regions. Expression of this antisense RNA is subject to each of the parameters described above in relation to expression of the sense RNA (regulatory sequences, constitutive or inducible expression, tissue-specific expression).

The invention also relates to recombinant host cells containing the vectors described herein. Host cells therefore include prokaryotic cells, lower eukaryotic cells such as yeast, other eukaryotic cells such as insect cells, and higher eukaryotic cells such as mammalian cells.

The recombinant host cells are prepared by introducing the vector constructs described herein into the cells by techniques readily available to the person of ordinary skill in the art. These include, but are not limited to, calcium phosphate transfection, DEAE-dextran-mediated transfection, cationic lipid-mediated transfection, electroporation, transduction, infection, lipofection, and other techniques such as those found in Sambrook, *et al.* (*Molecular Cloning: A Laboratory Manual. 2nd, ed., Cold Spring Harbor Laboratory,* Cold Spring Harbor Laboratory Press, Cold Spring Harbor, NY, 1989).

Host cells can contain more than one vector. Thus, different nucleotide sequences can be introduced on different vectors of the same cell. Similarly, the nucleic acid molecules can be introduced either alone or with other nucleic acid molecules that are not related to the nucleic acid molecules such as those providing trans-acting factors for expression vectors. When more than one vector is introduced into a cell, the vectors can be introduced independently, co-introduced or joined to the nucleic acid molecule vector.

In the case of bacteriophage and viral vectors, these can be introduced into cells as packaged or encapsulated virus by standard procedures for infection and transduction. Viral vectors can be replication-competent or replication-defective. In the case in which viral replication is defective, replication will occur in host cells providing functions that complement the defects.

Vectors generally include selectable markers that enable the selection of the subpopulation of cells that contain the recombinant vector constructs. The marker can be contained in the same vector that contains the nucleic acid molecules described herein or may be on a separate vector. Markers include tetracycline or ampicillin-resistance genes for prokaryotic host cells and dihydrofolate reductase or neomycin resistance for eukaryotic host cells. However, any marker that provides selection for a phenotypic trait will be effective.

While the mature proteins can be produced in bacteria, yeast, mammalian cells, and other cells under the control of the appropriate regulatory sequences, cell- free transcription and translation systems can also be used to produce these proteins using RNA derived from the DNA constructs described herein.

Where secretion of the peptide is desired, which is difficult to achieve with multi-transmembrane domain containing proteins such as GPCRs, appropriate secretion signals are

incorporated into the vector. The signal sequence can be endogenous to the peptides or heterologous to these peptides.

Where the peptide is not secreted into the medium, which is typically the case with GPCRs, the protein can be isolated from the host cell by standard disruption procedures, including freeze thaw, sonication, mechanical disruption, use of lysing agents and the like. The peptide can then be recovered and purified by well-known purification methods including ammonium sulfate precipitation, acid extraction, anion or cationic exchange chromatography, phosphocellulose chromatography, hydrophobic-interaction chromatography, affinity chromatography, hydroxylapatite chromatography, lectin chromatography, or high performance liquid chromatography.

It is also understood that depending upon the host cell in recombinant production of the peptides described herein, the peptides can have various glycosylation patterns, depending upon the cell, or maybe non-glycosylated as when produced in bacteria. In addition, the peptides may include an initial modified methionine in some cases as a result of a host-mediated process.

Uses of vectors and host cells

The recombinant host cells expressing the peptides described herein have a variety of uses. First, the cells are useful for producing a GPCR protein or peptide that can be further purified to produce desired amounts of GPCR protein or fragments. Thus, host cells containing expression vectors are useful for peptide production.

Host cells are also useful for conducting cell-based assays involving the GPCR protein or GPCR protein fragments, such as those described above as well as other formats known in the art. Thus, a recombinant host cell expressing a native GPCR protein is useful for assaying compounds that stimulate or inhibit GPCR protein function.

Host cells are also useful for identifying GPCR protein mutants in which these functions are affected. If the mutants naturally occur and give rise to a pathology, host cells containing the mutations are useful to assay compounds that have a desired effect on the mutant GPCR protein (for example, stimulating or inhibiting function) which may not be indicated by their effect on the native GPCR protein.

Genetically engineered host cells can be further used to produce non-human transgenic animals. A transgenic animal is preferably a mammal, for example a rodent, such as a rat or mouse, in which one or more of the cells of the animal include a transgene. A transgene is exogenous DNA which is integrated into the genome of a cell from which a transgenic animal

develops and which remains in the genome of the mature animal in one or more cell types or tissues of the transgenic animal. These animals are useful for studying the function of a GPCR protein and identifying and evaluating modulators of GPCR protein activity. Other examples of transgenic animals include non-human primates, sheep, dogs, cows, goats, chickens, and amphibians.

A transgenic animal can be produced by introducing nucleic acid into the male pronuclei of a fertilized oocyte, e.g., by microinjection, retroviral infection, and allowing the oocyte to develop in a pseudopregnant female foster animal. Any of the GPCR protein nucleotide sequences can be introduced as a transgene into the genome of a non-human animal, such as a mouse.

Any of the regulatory or other sequences useful in expression vectors can form part of the transgenic sequence. This includes intronic sequences and polyadenylation signals, if not already included. A tissue-specific regulatory sequence(s) can be operably linked to the transgene to direct expression of the GPCR protein to particular cells.

Methods for generating transgenic animals via embryo manipulation and microinjection, particularly animals such as mice, have become conventional in the art and are described, for example, in U.S. Patent Nos. 4,736,866 and 4,870,009, both by Leder *et al.*, U.S. Patent No. 4,873,191 by Wagner *et al.* and in Hogan, B., *Manipulating the Mouse Embryo*, (Cold Spring Harbor Laboratory Press, Cold Spring Harbor, N.Y., 1986). Similar methods are used for production of other transgenic animals. A transgenic founder animal can be identified based upon the presence of the transgene in its genome and/or expression of transgenic mRNA in tissues or cells of the animals. A transgenic founder animal can then be used to breed additional animals carrying the transgene. Moreover, transgenic animals carrying a transgene can further be bred to other transgenic animals carrying other transgenes. A transgenic animal also includes animals in which the entire animal or tissues in the animal have been produced using the homologously recombinant host cells described herein.

In another embodiment, transgenic non-human animals can be produced which contain selected systems that allow for regulated expression of the transgene. One example of such a system is the *cre/loxP* recombinase system of bacteriophage P1. For a description of the *cre/loxP* recombinase system, see, e.g., Lakso *et al. PNAS 89*:6232-6236 (1992). Another example of a recombinase system is the FLP recombinase system of *S. cerevisiae* (O'Gorman *et al. Science 251*:1351-1355 (1991). If a *cre/loxP* recombinase system is used to regulate expression of the transgene, animals containing transgenes encoding both the *Cre* recombinase and a selected protein is required. Such animals can be provided through the construction of

"double" transgenic animals, e.g., by mating two transgenic animals, one containing a transgene encoding a selected protein and the other containing a transgene encoding a recombinase.

Clones of the non-human transgenic animals described herein can also be produced according to the methods described in Wilmut, I. *et al. Nature 385*:810-813 (1997) and PCT International Publication Nos. WO 97/07668 and WO 97/07669. In brief, a cell, e.g., a somatic cell, from the transgenic animal can be isolated and induced to exit the growth cycle and enter G_o phase. The quiescent cell can then be fused, e.g., through the use of electrical pulses, to an enucleated oocyte from an animal of the same species from which the quiescent cell is isolated. The reconstructed oocyte is then cultured such that it develops to morula or blastocyst and then transferred to pseudopregnant female foster animal. The offspring born of this female foster animal will be a clone of the animal from which the cell, e.g., the somatic cell, is isolated.

Transgenic animals containing recombinant cells that express the peptides described herein are useful to conduct the assays described herein in an *in vivo* context. Accordingly, the various physiological factors that are present *in vivo* and that could effect ligand binding, GPCR protein activation, and signal transduction, may not be evident from *in vitro* cell-free or cell-based assays. Accordingly, it is useful to provide non-human transgenic animals to assay *in vivo* GPCR protein function, including ligand interaction, the effect of specific mutant GPCR proteins on GPCR protein function and ligand interaction, and the effect of chimeric GPCR proteins. It is also possible to assess the effect of null mutations, that is mutations that substantially or completely eliminate one or more GPCR protein functions.

All publications and patents mentioned in the above specification are herein incorporated by reference. Various modifications and variations of the described method and system of the invention will be apparent to those skilled in the art without departing from the scope and spirit of the invention. Although the invention has been described in connection with specific preferred embodiments, it should be understood that the invention as claimed should not be unduly limited to such specific embodiments. Indeed, various modifications of the above-described modes for carrying out the invention which are obvious to those skilled in the field of molecular biology or related fields are intended to be within the scope of the following claims.

Claims

That which is claimed is:

1. An isolated peptide consisting of an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

2. An isolated peptide comprising an amino acid sequence selected from the group consisting of:

(a) an amino acid sequence shown in SEQ ID NO:2;

(b) an amino acid sequence of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said allelic variant is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(c) an amino acid sequence of an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said ortholog is encoded by a nucleic acid molecule that hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic); and

(d) a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids.

3. An isolated antibody that selectively binds to a peptide of claim 2.

4. An isolated nucleic acid molecule consisting of a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

5. An isolated nucleic acid molecule comprising a nucleotide sequence selected from the group consisting of:

(a) a nucleotide sequence that encodes an amino acid sequence shown in SEQ ID NO:2;

(b) a nucleotide sequence that encodes of an allelic variant of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(c) a nucleotide sequence that encodes an ortholog of an amino acid sequence shown in SEQ ID NO:2, wherein said nucleotide sequence hybridizes under stringent conditions to the opposite strand of a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic);

(d) a nucleotide sequence that encodes a fragment of an amino acid sequence shown in SEQ ID NO:2, wherein said fragment comprises at least 10 contiguous amino acids; and

(e) a nucleotide sequence that is the complement of a nucleotide sequence of (a)-(d).

6. A gene chip comprising a nucleic acid molecule of claim 5.

7. A transgenic non-human animal comprising a nucleic acid molecule of claim 5.

8. A nucleic acid vector comprising a nucleic acid molecule of claim 5.

9. A host cell containing the vector of claim 8.

10. A method for producing any of the peptides of claim 1 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

11. A method for producing any of the peptides of claim 2 comprising introducing a nucleotide sequence encoding any of the amino acid sequences in (a)-(d) into a host cell, and culturing the host cell under conditions in which the peptides are expressed from the nucleotide sequence.

12. A method for detecting the presence of any of the peptides of claim 2 in a sample, said method comprising contacting said sample with a detection agent that specifically allows detection of the presence of the peptide in the sample and then detecting the presence of the peptide.

13. A method for detecting the presence of a nucleic acid molecule of claim 5 in a sample, said method comprising contacting the sample with an oligonucleotide that hybridizes to said nucleic acid molecule under stringent conditions and determining whether the oligonucleotide binds to said nucleic acid molecule in the sample.

14. A method for identifying a modulator of a peptide of claim 2, said method comprising contacting said peptide with an agent and determining if said agent has modulated the function or activity of said peptide.

15. The method of claim 14, wherein said agent is administered to a host cell comprising an expression vector that expresses said peptide.

16. A method for identifying an agent that binds to any of the peptides of claim 2, said method comprising contacting the peptide with an agent and assaying the contacted mixture to determine whether a complex is formed with the agent bound to the peptide.

17. A pharmaceutical composition comprising an agent identified by the method of claim 16 and a pharmaceutically acceptable carrier therefor.

18. A method for treating a disease or condition mediated by a human proteases, said method comprising administering to a patient a pharmaceutically effective amount of an agent identified by the method of claim 16.

19. A method for identifying a modulator of the expression of a peptide of claim 2, said method comprising contacting a cell expressing said peptide with an agent, and determining if said agent has modulated the expression of said peptide.

20. An isolated human protease peptide having an amino acid sequence that shares at least 70% homology with an amino acid sequence shown in SEQ ID NO:2.

21. A peptide according to claim 20 that shares at least 90 percent homology with an amino acid sequence shown in SEQ ID NO:2.

22. An isolated nucleic acid molecule encoding a human protease peptide, said nucleic acid molecule sharing at least 80 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

23. A nucleic acid molecule according to claim 22 that shares at least 90 percent homology with a nucleic acid molecule shown in SEQ ID NOS:1 (transcript) or 3 (genomic).

```
   1  CCCGGGCTGC AGGAATTCGG CACGAGGCCA TGCTGGGCCC TGCTGTCCTG
  51  GGCCTCAGCC TCTGGGCTCT CCTGCACCCT GGGACGGGGG CCCCATTGTG
 101  CCTGTCACAG CAACTTAGGA TGAAGGGGGA CTACGTGCTG GGGGGGCTGT
 151  TCCCCCTGGG CGAGGCCGAG GAGGCTGGCC TCCGCAGCCG GACACGGCCC
 201  AGCAGCCCTG TGTGCACCAG GTTCTCCTCA AACGGCCTGC TCTGGGCACT
 251  GGCCATGAAA ATGGCCGTGG AGGAGATCAA CAACAAGTCG GATCTGCTGC
 301  CCGGGCTGCG CCTGGGCTAC GACCTCTTTG ATACGTGCTC GGAGCCTGTG
 351  GTGGCCATGA AGCCCAGCCT CATGTTCCTG GCCAAGGCAG GCAGCCGCGA
 401  CATCGCCGCC TACTGCAACT ACACGCAGTA CCAGCCCCGT GTGCTGGCTG
 451  TCATCGGGCC CCACTCGTCA GAGCTCGCCA TGGTCACCGG CAAGTTCTTC
 501  AGCTTCTTCC TCATGCCCCA GGTCAGCTAC GGTGCTAGCA TGGAGCTGCT
 551  GAGCGCCCGG GAGACCTTCC CCTCCTTCTT CCGCACCGTG CCCAGCGACC
 601  GTGTGCAGCT GACGGCCGCC GCGGAGCTGC TGCAGGAGTT CGGCTGGAAC
 651  TGGGTGGCCG CCCTGGGCAG CGACGACGAG TACGGCCGGC AGGGCCTGAG
 701  CATCTTCTCG GCCCTGGCCG CGGCACGCGG CATCTGCATC GCGCACGAGG
 751  GCCTGGTGCC GCTGCCCCGT GCCGATGACT CGCGGCTGGG GAAGGTGCAG
 801  GACGTCCTGC ACCAGGTGAA CCAGAGCAGC GTGCAGGTGG TGCTGCTGTT
 851  CGCCTCCGTG CACGCCGCCC ACGCCCTCTT CAACTACAGC ATCAGCAGCA
 901  GGCTCTCGCC CAAGGTGTGG GTGGCCAGCG AGGCCTGGCT GACCTCTGAC
 951  CTGGTCATGG GGCTGCCCGG CATGGCCCAG ATGGGCACGG TGCTTGGCTT
1001  CCTCCAGAGG GGTGCCCAGC TGCACGAGTT CCCCCAGTAC GTGAAGACGC
1051  ACCTGGCCCT GGCCACCGAC CCGGCCTTCT GCTCTGCCCT GGGCGAGAGG
1101  GAGCAGGGTC TGGAGGAGGA CGTGGTGGGC CAGCGCTGCC CGCAGTGTGA
1151  CTGCATCACG CTGCAGAACG TGAGCGCAGG GCTAAATCAC CACCAGACGT
1201  TCTCTGTCTA CGCAGCTGTG TATAGCGTGG CCCAGGCCCT GCACAACACT
1251  CTTCAGTGCA ACGCCTCAGG CTGCCCCGCG CAGGACCCCG TGAAGCCCTG
1301  GCAGCTCCTG GAGAACATGT ACAACCTGAC CTTCCACGTG GGCGGGCTGC
1351  CGCTGCGGTT CGACAGCAGC GGAAACGTGG ACATGGAGTA CGACCTGAAG
1401  CTGTGGGTGT GGCAGGGCTC AGTGCCCAGG CTCCACGACG TGGGCAGGTT
1451  CAACGGCAGC CTCAGGACAG AGCGCCTGAA GATCCGCTGG CACACGTCTG
1501  ACAACCAGAA GCCCGTGTCC CGGTGCTCGC GGCAGTGCCA GGAGGGCCAG
1551  GTGCGCCGGG TCAAGGGGTT CCACTCCTGC TGCTACGACT GTGTGGACTG
1601  CGAGGCGGGC AGCTACCGGC AAAACCCAGA CGACATCGCC TGCACCTTTT
1651  GTGGCCAGGA TGAGTGGTCC CCGGAGCGAA GCACACGCTG CTTCCGCCGC
1701  AGGTCTCGGT TCCTGGCATG GGGCGAGCCG GCTGTGCTGC TGCTGCTCCT
1751  GCTGCTGAGC CTGGCGCTGG GCCTTGTGCT GGCTGCTTTG GGGCTGTTCG
1801  TTCACCATCG GGACAGCCCA CTGGTTCAGG CCTCGGGGGG GCCCCTGGCC
1851  TGCTTTGGCC TGGTGTGCCT GGGCCTGGTC TGCCTCAGCG TCCTCCTGTT
1901  CCCTGGCCAG CCCAGCCCTG CCCGATGCCT GGCCCAGCAG CCCTTGTCCC
1951  ACCTCCCGCT CACGGGCTGC CTGAGCACAC TCTTCCTGCA GGCGGCCGAG
2001  ATCTTCGTGG AGTCAGAACT GCCTCTGAGC TGGGCAGACC GGCTGAGTGG
2051  CTGCCTGCGG GGGCCCTGGG CCTGGCTGGT GGTGCTGCTG GCCATGCTGG
2101  TGGAGGTCGC ACTGTGCACC TGGTACCTGG TGGCCTTCCC GCCGGAGGTG
2151  GTGACGGACT GGCACATGCT GCCCACGGAG GCGCTGGTGC ACTGCCGCAC
2201  ACGCTCCTGG GTCAGCTTCG GCCTAGCGCA CGCCACCAAT GCCACGCTGG
2251  CCTTTCTCTG CTTCCTGGGC ACTTTCCTGG TGCGGAGCCA GCCGGGCCGC
2301  TACAACCGTG CCCGTGGCCT CACCTTTGCC ATGCTGGCCT ACTTCATCAC
2351  CTGGGTCTCC TTTGTGCCCC TCCTGGCCAA TGTGCAGGTG GTCCTCAGGC
2401  CCGCCGTGCA GATGGGCGCC CTCCTGCTCT GTGTCCTGGG CATCCTGGCT
2451  GCCTTCCACC TGCCCAGGTG TTACCTGCTC ATGCGGCAGC CAGGGCTCAA
2501  CACCCCCGAG TTCTTCCTGG GAGGGGGCCC TGGGGATGCC CAAGGCCAGA
2551  ATGACGGGAA CACAGGAAAT CAGGGGAAAC ATGAGTGACC CAACCCTGTG
2601  ATCTCAGCCC CGGTGAACCC AGACTTAGCT GCGATCCCCC CCAAGCCAGC
2651  AATGACCCGT GTCTCGCTAC AGAGACCCTC CCGCTCTAGG TTCTGACCCC
2701  AGGTTGTCTC CTGACCCTGA CCCCACAGTG AGCCCTAGGC CTGGAGCACG
2751  TGGACACCCC TGTGACCATC TGGGCCCCAG AGCCAAGCTG TGTCCCTGTC
2801  CCTCTGTGCC CAGACCAGGC CTGCCCAGGT AACCCAGACC CACTGTTCTG
2851  GAAAGAGGCC CGGAGGGCTC CCAGGGTACC CGCAACCCAC ACCGTGAGCT
2901  CAGGAAAAGG ACGCAGGGAG GCCCCGGCCA GATGGCTGGA AGCCCAAATC
2951  AGGCCCTGCC GACCTGACCA TGTCCCACCA GGGCCCCCAT CCTGCACCCT
3001  GCCAGGCACC ACAGCAGTGG GAGGCCAGGT GGGGGCACAC AGGCATATGC
3051  CCAGGGCAGA GCCCGCCGAG GTGGGGGTGG CACCCAGCTT CCTACTCTGC
3101  CCCTTGCCCA GTGGGTAGAC AGCATCATGA CTGTCACCAG TACCAGGGAC
3151  AGAGCCCAGG TGGGGTGGGG GCGGGGTCCA GCACCACGGC CAGCACCGAC
3201  CACCAGGACC CCGGAGCCAG CACCATGGAC AGAAAACTGC CCACCAGGAT
3251  CTGACGCCAG CACGCCGCCA GGCCCACACA GGGTCTCCGG TCAGAGTCCC
3301  AGGGTCAGCT CCCAGCAGGG CCTAGGGGAG GCTGGACCAG CTCCCTGTGC
```

3351 CTCATTCCAA GGCAGCCCAG CCGGAGAGAA GGGGCACAGG CCACACATCT
3401 GTCCCATAAA ATTAAACGCT TTTTAGTGTT TAAAATAAAA AAAAAAAAAA
3451 AAAAAAAA
(SEQ ID NO:1)

FEATURES:
Start: 30
Stop: 2586

HOMOLOGOUS PROTEIN:
Top 10 BLAST Hits:

	Score	E
gi|4337086|gb|AAD18069.1| (AF127389) putative taste receptor TR...	460	e-128
gi|1168781|sp|P41180|CASR_HUMAN EXTRACELLULAR CALCIUM-SENSING R...	395	e-109
gi|4557411|ref|NP_000379.1| calcium-sensing receptor (hypocalci...	395	e-109
gi|904210|dbj|BAA09453.1| (D50855) Ca-sensing receptor [Homo sa...	395	e-109
gi|1082270|pir||S49341 calcium-sensing receptor - human	395	e-109
gi|7305101|ref|NP_038831.1| G protein coupled receptor, family ...	394	e-108
gi|543933|sp|P35384|CASR_BOVIN EXTRACELLULAR CALCIUM-SENSING RE...	392	e-108
gi|8393053|ref|NP_058692.1| calcium-sensing receptor (hypocalci...	391	e-107
gi|5163340|gb|AAD40638.1|AF128842_1 (AF128842) extracellular ca...	391	e-107
gi|1362762|pir||B56715 calcium receptor (clone phPCaR-5.2) - hu...	389	e-107

EST:

	Score	E
gi|3042482|gb|AA907022.1|AA907022 oj92a08.s2 Soares_NFL_T_GBC_S...	198	7e-48

EXPRESSION INFORMATION FOR MODULATORY USE:
library source:
Expression information from BLAST EST hit:
gi|3042482|gb|AA907022.1|AA907022 oj92a08.s2 fetal lung NbHL19W, testis NHT, and B-cell (pooled)

Expression information from cDNA library screening:
Human Hela cells
Human Bone Marrow

```
      1  MLGPAVLGLS LWALLHPGTG APLCLSQQLR MKGDYVLGGL FPLGEAEEAG
     51  LRSRTRPSSP VCTRFSSNGL LWALAMKMAV EEINNKSDLL PGLRLGYDLF
    101  DTCSEPVVAM KPSLMFLAKA GSRDIAAYCN YTQYQPRVLA VIGPHSSELA
    151  MVTGKFFSFF LMPQVSYGAS MELLSARETF PSFFRTVPSD RVQLTAAAEL
    201  LQEFGWNWVA ALGSDDEYGR QGLSIFSALA AARGICIAHE GLVPLPRADD
    251  SRLGKVQDVL HQVNQSSVQV VLLFASVHAA HALFNYSISS RLSPKVWVAS
    301  EAWLTSDLVM GLPGMAQMGT VLGFLQRGAQ LHEFPQYVKT HLALATDPAF
    351  CSALGEREQG LEEDVVGQRC PQCDCITLQN VSAGLNHHQT FSVYAAVYSV
    401  AQALHNTLQC NASGCPAQDP VKPWQLLENM YNLTFHVGGL PLRFDSSGNV
    451  DMEYDLKLWV WQGSVPRLHD VGRFNGSLRT ERLKIRWHTS DNQKPVSRCS
    501  RQCQEGQVRR VKGFHSCCYD CVDCEAGSYR QNPDDIACTF CGQDEWSPER
    551  STRCFRRRSR FLAWGEPAVL LLLLLLSLAL GLVLAALGLF VHHRDSPLVQ
    601  ASGGPLACFG LVCLGLVCLS VLLFPGQPSP ARCLAQQPLS HLPLTGCLST
    651  LFLQAAEIFV ESELPLSWAD RLSGCLRGPW AWLVVLLAML VEVALCTWYL
    701  VAFPPEVVTD WHMLPTEALV HCRTRSWVSF GLAHATNATL AFLCFLGTFL
    751  VRSQPGRYNR ARGLTFAMLA YFITWVSFVP LLANVQVVLR PAVQMGALLL
    801  CVLGILAAFH LPRCYLLMRQ PGLNTPEFFL GGGPGDAQGQ NDGNTGNQGK
    851  HE
(SEQ ID NO:2)
```

FEATURES:
Functional domains and key regions:

```
[1] PDOC00001 PS00001 ASN_GLYCOSYLATIONN-glycosylation site
Number of matches: 9
      1      85-88 NKSD
      2    130-133 NYTQ
      3    264-267 NQSS
      4    285-288 NYSI
      5    380-383 NVSA
      6    411-414 NASG
      7    432-435 NLTF
      8    475-478 NGSL
      9    737-740 NATL
-------------------------------------------------------------------------------[2]
PDOC00004 PS00004 CAMP_PHOSPHO_SITEcAMP- and cGMP-dependent protein kinase
phosphorylation site
          556-559 RRRS
-------------------------------------------------------------------------------[3]
PDOC00005 PS00005 PKC_PHOSPHO_SITEProtein kinase C phosphorylation site
Number of matches: 9
      1    153-155 TGK
      2    175-177 SAR
      3    189-191 SDR
      4    289-291 SSR
      5    293-295 SPK
      6    477-479 SLR
      7    480-482 TER
      8    528-530 SYR
      9    551-553 STR
-------------------------------------------------------------------------------[4]
PDOC00006 PS00006 CK2_PHOSPHO_SITECasein kinase II phosphorylation site
Number of matches: 4
      1    102-105 TCSE
      2    175-178 SARE
      3    214-217 SDDE
      4    667-670 SWAD
-------------------------------------------------------------------------------[5]
```

PDOC00008 PS00008 MYRISTYLN-myristoylation site
Number of matches: 15

1	20-25	GAPLCL
2	69-74	GLLWAL
3	92-97	GLRLGY
4	234-239	GICIAH
5	319-324	GTVLGF
6	476-481	GSLRTE
7	581-586	GLVLAA
8	603-608	GGPLAC
9	646-651	GCLSTL
10	731-736	GLAHAT
11	763-768	GLTFAM
12	804-809	GILAAF
13	831-836	GGGPGD
14	839-844	GQNDGN
15	843-848	GNTGNQ

----------[6]

PDOC00754 PS00980 G_PROTEIN_RECEP_F3_2G-protein coupled receptors family 3 signature 2
517-541 CCYDCVDCEAGSYRQNPDDIACTFC

Membrane spanning structure and domains:

Helix	Begin	End	Score	Certainty
1	137	167	1.179	Certain
2	221	241	0.682	Putative
3	256	286	1.149	Certain
4	306	326	1.363	Certain
5	388	408	0.700	Putative
6	567	587	2.340	Certain
7	606	626	2.253	Certain
8	637	657	1.176	Certain
9	680	700	1.695	Certain
10	731	751	1.543	Certain
11	763	783	2.171	Certain
12	792	812	1.955	Certain

```
BLAST Alignment to Top Hit:
>gi|4337086|gb|AAD18069.1| (AF127389) putative taste receptor TR1
           [Rattus norvegicus]
           Length = 840

 Score =  460 bits (1170), Expect = e-128
 Identities = 276/816 (33%), Positives = 425/816 (51%), Gaps = 39/816 (4%)

Query: 31  MKGDYVLGGLFPL-GEAEEAGLRSRTRPSSPVCTR---FSSNGLLWALAMKMAVEEINNK 86
           + GD++L GLF L G+     L+ R RP    C R   F+ +G     AM+  VEEINN
Sbjct: 33  LPGDFLLAGLFSLHGDC----LQVRHRPLVTSCDRPDSFNGHGYHLFQAMRFTVEEINNS 88

Query: 87  SDLLPGLRLGYDLFDTCSEPVVAMKPSLMFLAKAGSRDIAAYCNYTQYQPRVLAVIGPHS 146
           S LLP + LGY+L+D CSE    +  +L  LA  G R I    +   +  +V+A IGP +
Sbjct: 89  SALLPNITLGYELYDVCSESA-NVYATLRVLALQGPRHIEIQKDLRNHSSKVVAFIGPDN 147

Query: 147 SELAMVTGKFFSFFLMPQVSYGASMELLSARETFPSFFRTVPSDRVQLTAAAELLQEFGW 206
           ++ A+ T       FLMP VSY AS  +LSA+  FPSF RTVPSDR Q+    +LLQ FGW
Sbjct: 148 TDHAVTTAALLGPFLMPLVSYEASSVVLSAKRKFPSFLRTVPSDRHQVEVMVQLLQSFGW 207

Query: 207 NWVAALGSDDEYGRQGLSIFSALAAARGICIAHEGLVPLP-RADDSRLGKVQDVLHQVNQ 265
            W++ +GS  +YG+ G+     LA  RGIC+A + +VP   R  D R+   Q ++  + Q
Sbjct: 208 VWISLIGSYGDYGQLGVQALEELAVPRGICVAFKDIVPFSARVGDPRM---QSMMQHLAQ 264

Query: 266 SSVQVVLLFASVHAAHALFNYSISSRLSPKVWVASEAWLTSDLVMGLPGMAQMGTVLGFL 325
           +   VV++F++ H A   F   + + L+ KVWVASE W  S  +  + G+  +GTVLG
Sbjct: 265 ARTTVVVVFSNRHLARVFFRSVVLANLTGKVWVASEDWAISTYITSVTGIQGIGTVLGVA 324

Query: 326 QRGAQLHEFPQYVKTHLALAT--------DPAFCSALGEREQGLEEDVVGQRCPQCDCITL 378
            +  Q+    ++ ++++   T        + ++CS                Q C +C    T
Sbjct: 325 VQQRQVPGLKEFEESYVRAVTAAPSACPEGSWCST-------------NQLCRECHTFTT 371

Query: 379 QNVSA--GLNHHQTFSVYAAVYSVAQALHNTLQCNASGCPAQDPVKPWQLLENMYNLTFH 436
           +N+      +    + VY AVY+VA  LH  L C +  C ++ PV PWQLL+ +Y + F
Sbjct: 372 RNMPTLGAFSMSAAYRVYEAVYAVAHGLHQLLGCTSEIC-SRGPVYPWQLLQQIYKVNFL 430

Query: 437 VGGLPLRFDSSGNVDMEYDLKLWVWQGSVPRLHDVGRFNGS---LRTERLKIRWHTSDNQ 493
           +    + FD +G+    YD+  W W G       +G  + S   L   + KI+WH  +NQ
Sbjct: 431 LHENTVAFDDNGDTLGYYDIIAWDWNGPEWTFEIIGSASLSPVHLDINKTKIQWHGKNNQ 490

Query: 494 KPVSRCSRQCQEGQVRRVKGFHSCCYDCVDCEAGSYRQNPDDIACTFCGQDEWSPERSTR 553
            PVS C+  C  G  R V G H CC++CV CEAG++    +   C  CG +EW+P+ ST
Sbjct: 491 VPVSVCTTDCLAGHHRVVVGSHHCCFECVPCEAGTFLNMSELHICQPCGTEEWAPKESTT 550

Query: 554 CFRRRSRFLAWGEPAVLLLLLLLSLALGLVLAALGLFVHHRDSPLVQASGGPLACFGLVC 613
           CF R   FLAW EP  L+L+   +L L L++   GLF  H  +P+V+++GG L    L
Sbjct: 551 CFPRTVEFLAWHEPISLVLIAANTLLLLLLVGTAGLFAWHFHTPVVRSAGGRLCFLMLGS 610

Query: 614 LGLVCLSVLLFPGQPSPARCLAQQPLSHLPLTGCLSTLFLQAAEIFVESELPLSWADRLS 673
           L     S   F G+P+   CL +QPL  L     LS L +++ ++ +  +
Sbjct: 611 LVAGSCSFYSFFGEPTVPACLLRQPLFSLGFAIFLSCLTIRSFQLVIIFKFSTKVPTFYR 670

Query: 674 GCLRGPWAWLVVLLAMLVEVALCTWYLVAFPPEVVTDWHMLPTEALVHCRTRSWVSFGLA 733
              + A L V+++  V + +C  +LV + P    ++   P   ++ C   + V F LA
Sbjct: 671 TWAQNHGAGLFVIVSSTVHLLICLTWLVMWTPRPTREYQRFPHLVILECTEVNSVGFLLA 730

Query: 734 HATNATLAFLCFLGTFLVRSQPGRYNRARGLTFAMLAYFITWVSFVPLLANVQVVLRPAV 793
              N  L+   F+ ++L +  P  YN A+ +TF++L  F++W++F  + +  Q    PAV
Sbjct: 731 FTHNILLSISTFVCSYLGKELPENYNEAKCVTFSLLLNFVSWIAFFTMASIYQGSYLPAV 790
```

```
Query: 794 QMGALLLCVLGILAAFHLPRCYLLMRQPGLNTPEFF 829
          + A L  + G  + + LP+CY+++ +P LN  E F
Sbjct: 791 NVLAGLTTLSGGFSGYFLPKCYVILCRPELNNTEHF 826
(SEQ ID NO:4)
```

Hmmer search results (Pfam):

Model	Description	Score	E-value	N
PF01094	Receptor family ligand binding region	127.0	9.8e-35	2
CE00344	E00344 parathyroid_cell_calcium-sensing_rece	70.5	2.2e-21	4
CE00165	CE00165 Metabotropic_glutamate_receptor	58.7	2.1e-18	4
CE00294	E00294 glutamate_receptor_6	26.1	1.4e-07	2
PF00003	7 transmembrane receptor (metabotropic gluta	12.1	0.012	2
PF01796	Domain of unknown function	6.0	2.1	1
PF01160	Vertebrate endogenous opioids neuropeptide	3.6	4.6	1

Parsed for domains:

Model	Domain	seq-f	seq-t		hmm-f	hmm-t		score	E-value
CE00165	1/4	33	42	..	1	10	[.	5.8	0.13
CE00344	1/4	32	103	..	30	102	..	17.1	4.3e-05
CE00165	2/4	75	104	..	39	68	..	9.6	0.008
CE00344	2/4	160	237	..	161	238	..	37.4	2.7e-11
CE00294	1/2	75	238	..	71	264	..	25.7	1.8e-07
CE00165	3/4	183	239	..	159	216	..	29.3	4.5e-09
PF01094	1/2	61	276	..	1	232	[.	98.3	1.1e-26
PF01796	1/1	352	373	..	1	23	[.	6.0	2.1
PF01094	2/2	393	456	..	362	430	..	28.7	3.3e-07
CE00344	3/4	517	567	..	562	612	..	6.2	0.089
CE00294	2/2	396	603	..	431	649	..	1.3	3.4
CE00165	4/4	484	603	..	481	599	..	13.7	0.00041
PF00003	1/2	566	639	..	1	72	[.	5.0	1.4
PF01160	1/1	691	709	..	1	19	[.	3.6	4.6
CE00344	4/4	740	780	..	784	824	..	9.8	0.007
PF00003	2/2	740	824	..	180	270	.]	8.3	0.15

```
   1  GCAAAAGAGG AGAGGCCTGC GTGGGATGTG TGCGTGGAGG TGGGGGCCAC
  51  TCCCAGCCAC AACAGTGGCC TGGACAGAAA GGGGATGCAA TTCAGCAGAG
 101  TCTTCCTGGA TGTCACCCCC ACCTCAGGGT CTGTGGGAGC TGCATATGGT
 151  GCCGGCAAAG GCTGCCGACT GCAGTATGGG CCGGGAGAAC TGCCTGGGTC
 201  TGCGTGGGCC CCAGGGCAGG GCTCCCTCCG GGTGTTGCCT TCTGTACAAG
 251  TGCCATGCTT GTGCCGTTTG CGTGTCCCAA GTGCGAGTGT GCTATTTGCG
 301  TGTGCCGCAC GTGTGCCGTT TGCATGTGCT GTTTGCATGT ACCATGTGCA
 351  TGTGTGCCAT CTGTGCAATG TGCAGGTGCC AGTTGCATGT GCCATGCGTG
 401  TTGGCTGTGA GCGTGTGCTG TTTTCGTGTA TGTGCCATGC ACGTATGTGC
 451  TGCGTGTTGG CCGTGCACGT GTGCCACGTG CATGTGTGCC ATTTGTGTAT
 501  GCCGTGTGTG CTGTGAGCGT ATGCTGTGCG CATGTGCGTG CCATGCGCTT
 551  TGCGTGTACC ATGTGTGTGC TGTTTGCATG TGCCATTTGT GTCATGCACG
 601  TGCCATTTGC GTGCGATGCG CGTGTGCCAC ACGTCGTTTG CTTGCGTGCC
 651  ATGCATGTGT GGCATTTGTG TATGTGCCGT TTCCGTGTGT ATTGTGTGTG
 701  CCGTGTGTGT GCCATTTGCA TGTGCCGTTT TGTGTGTGCC ATGCGCGTGT
 751  GCCATGCACT TGCCGTGCGT GTGCCATTTG TGTGTACCAT GCGCATGTGC
 801  CATTCGTGTG CACCGTACAC GTGTGCCATT TGCATGTATG CTGTGCACGT
 851  GCGGCATGCA TGTGTGCCGT TTGCATGCCA TGCATGTGTT CCTTGCGTGT
 901  GCCGTGCGTG TCCCATGCAC GTGTGCCGTG CATGTGCCAT TCGCGTGTAC
 951  CATGCGCATG TGCCGTTTGT GTGTGTGCCG CATTCCTGTG CTCGTGTGCC
1001  AGGTTCGCAT GTGCGCCATA TTCACGTGTG CTCAGCATGT GCCATATGCA
1051  TGTGCGGTGG TAGTGTGTGT CCCTCACAGG TCCTCCTCAC AACACCATGG
1101  GGAAGAAGCA CCAGCCAGGG CACACCTCCT GGTATCTGCT AGGTCTGCCA
1151  GGCCCTAGCT GAAGCTGAGT GCCCCCCAGT TCCCCTGGGA GGGCCTGCGC
1201  CTGGAGTCTG CTGTGTCCCC GAGGGCACCC CCAAAGCAAC ACAGAGGCAG
1251  AGGAGTCCCG GCCCTGCACA CCTGGTGCTG CTCCAGCTGC CGCTCATTTG
1301  CCTGTGGCCC TTCCTCCCTT GTTTGCGTGC CCCCCTGGCA AACAAACTCT
1351  ACCCCCAGCA GGAGCCACCT GTGTGCCTGC CACGCAGGAG TGGCCCAGAC
1401  GGGGGTCAGC AGTGTGAGTA CAGCTGGCCA TGCGGTTCCT ACAGCTTCCA
1451  GGCGTCAGAC TCTGGCAGAA GGGCTGAGAC CCTCAAGGAA CTCTGCTCCC
1501  AAGCAGACTG GGAGGGCAGC ACCACCACCC CAGGGCCCTC CCCAGCTGCA
1551  GGGTGGAGGC CTGGCTGGCC GGCTGCCCAC TGGCCTGACT GGTCTGCAGG
1601  CCTAGGGGGC CCATCCCTGC TGCCCCCGGC TCCGGCCAGC ACAGCCTTGA
1651  GTGGGAGCCA GAAGCTCCCG GGGCTGGGTA GGAGGCATTT CTGTGCTTAT
1701  GAAAAGCCCC AGGGCTGGGT GTCTCTGCAT CCCTCCCACG CAGCTGAGAC
1751  CTCAGAGCCC TGGAGGCCCC TTTGCCCCCT CTCCTCTCCA CAGCCTGCTG
1801  GGCAACTCCA GGAATCGGGG GGTGGCAAGG GGCTCAGCCA CAGGCAGGGA
1851  ACAAGGCCAC GGCCAGCGAC TGAGCAGAGC CTGCCTGCCG GTCAACGCTG
1901  GCCATAGAGC CTGGCAGTGG CCTCAGGCAG AGTCTGACGC GCACAAACTT
1951  TCAGGCCCAG GAAGCGAGGA CACCACTGGG GCCCCAGGGT GTGGCAAGTG
2001  AGGATGGCAA GGGTTTTGCT AAACAAATCC TCTGCCCGCT CCCCGCCCCG
2051  GGCTCACTCC ATGTGAGGCC CCAGTCGGGG CAGCCACCTG CCGTGCCTGT
2101  TGGAAGTTGC CTCTGCCATG CTGGGCCCTG CTGTCCTGGG CCTCAGCCTC
2151  TGGGCTCTCC TGCACCCTGG GACGGGGGCC CCATTGTGCC TGTCACAGCA
2201  ACTTAGGATG AAGGGGGACT ACGTGCTGGG GGGGCTGTTC CCCCTGGGCG
2251  AGGCCGAGGA GGCTGGCCTC CGCAGCCGGA CACGGCCCAG CAGCCCTGTG
2301  TGCACCAGGT ACAGAGGTGG GACGGCCTGG GTCGGGGTCA GGGTGACCAG
2351  GTCTGGGGTG CTCCTGAGCT GGGGCCGAGG TGGCCATCTG CGGTTCTGTG
2401  TGGCCCCAGG TTCTCCTCAA ACGGCCTGCT CTGGGCACTG GCCATGAAAA
2451  TGGCCGTGGA GGAGATCAAC AACAAGTCGG ATCTGCTGCC CGGGCTGCGC
2501  CTGGGCTACG ACCTCTTTGA TACGTGCTCG GAGCCTGTGG TGGCCATGAA
2551  GCCCAGCCTC ATGTTCCTGG CCAAGGCAGG CAGCCGCGAC ATCGCCGCCT
2601  ACTGCAACTA CACGCAGTAC CAGCCCCGTG TGCTGGCTGT CATCGGGCCC
2651  CACTCGTCAG AGCTCGCCAT GGTCACCGGC AAGTTCTTCA GCTTCTTCCT
2701  CATGCCCCAG GTGGGCGCCC CCCACCATCA CCCACCCCCA CCCAGCCCTG
2751  CCCGTGGGAG CCCCTGTGTC AGGAGATGCC TCTTGGCCCT TGCAGGTCAG
2801  CTACGGTGCT AGCATGGAGC TGCTGAGCGC CCGGGAGACC TTCCCCTCCT
2851  TCTTCCGCAC CGTGCCCAGC GACCGTGTGC AGCTGACGGC CGCCGCGGAG
2901  CTGCTGCAGG AGTTCGGCTG GAACTGGGTG GCCGCCCTGG GCAGCGACGA
2951  CGAGTACGGC CGGCAGGGCC TGAGCATCTT CTCGGCCCTG GCCGCGGCAC
3001  GCGGCATCTG CATCGCGCAC GAGGGCCTGG TGCCGCTGCC CCGTGCCGAT
3051  GACTCGCGGC TGGGGAAGGT GCAGGACGTC CTGCACCAGG TGAACCAGAG
3101  CAGCGTGCAG GTGGTGCTGC TGTTCGCCTC CGTGCACGCC GCCCACGCCC
3151  TCTTCAACTA CAGCATCAGC AGCAGGCTCT CGCCCAAGGT GTGGGTGGCC
3201  AGCGAGGCCT GGCTGACCTC TGACCTGGTC ATGGGGCTGC CCGGCATGGC
3251  CCAGATGGGC ACGGTGCTTG GCTTCCTCCA GAGGGGTGCC CAGCTGCACG
3301  AGTTCCCCCA GTACGTGAAG ACGCACCTGG CCCTGGCCAC CGACCCGGCC
```

```
3351  TTCTGCTCTG CCCTGGGCGA GAGGGAGCAG GGTCTGGAGG AGGACGTGGT
3401  GGGCCAGCGC TGCCCGCAGT GTGACTGCAT CACGCTGCAG AACGTGAGCG
3451  CAGGGCTAAA TCACCACCAG ACGTTCTCTG TCTACGCAGC TGTGTATAGC
3501  GTGGCCCAGG CCCTGCACAA CACTCTTCAG TGCAACGCCT CAGGCTGCCC
3551  CGCGCAGGAC CCCGTGAAGC CCTGGCAGGT GAGCCCGGGA GATGGGGGTG
3601  TGCTGTCCTC TGCATGTGCC CAGGCCACCA GGCACGGCCA CCACGCCTGA
3651  GCTGGAGGTG GCTGGCGGCT CAGCCCCGTC CCCCGCCCGC AGCTCCTGGA
3701  GAACATGTAC AACCTGACCT TCCACGTGGG CGGGCTGCCG CTGCGGTTCG
3751  ACAGCAGCGG AAACGTGGAC ATGGAGTACG ACCTGAAGCT GTGGGTGTGG
3801  CAGGGCTCAG TGCCCAGGCT CCACGACGTG GGCAGGTTCA ACGGCAGCCT
3851  CAGGACAGAG CGCCTGAAGA TCCGCTGGCA CACGTCTGAC AACCAGGTGA
3901  GGTGAGGGTG GGTGTGCCAG GCGTGCCCGT GGTAGCCCCC GCGGCAGGGC
3951  GCAGCCTGGG GGTGGGGGCC GTTCCAGTCT CCCGTGGGCA TGCCCAGCCG
4001  AGCAGAGCCA GACCCCAGGC CTGTGCGCAG AAGCCCGTGT CCCGGTGCTC
4051  GCGGCAGTGC CAGGAGGGCC AGGTGCGCCG GGTCAAGGGG TTCCACTCCT
4101  GCTGCTACGA CTGTGTGGAC TGCGAGGCGG GCAGCTACCG GCAAAACCCA
4151  GGTGAGCCGC CTTCCCGGCA GGCGGGGGTG GGAACGCAGC AGGGGAGGGT
4201  CCTGCCAAGT CCTGACTCTG AGACCAGAGC CCACAGGGGA CAAGACGAAC
4251  ACCCAGCGCC CTTCTCCTCT CTCACAGACG ACATCGCCTG CACCTTTTGT
4301  GGCCAGGATG AGTGGTCCCC GGAGCGAAGC ACACGCTGCT TCCGCCGCAG
4351  GTCTCGGTTC CTGGCATGGG GCGAGCCGGC TGTGCTGCTG CTGCTCCTGC
4401  TGCTGAGCCT GGCGCTGGGC CTTGTGCTGG CTGCTTTGGG GCTGTTCGTT
4451  CACCATCGGG ACAGCCCACT GGTTCAGGCC TCGGGGGGGC CCCTGGCCTG
4501  CTTTGGCCTG GTGTGCCTGG GCCTGGTCTG CCTCAGCGTC CTCCTGTTCC
4551  CTGGCCAGCC CAGCCCTGCC CGATGCCTGG CCCAGCAGCC CTTGTCCCAC
4601  CTCCCGCTCA CGGGCTGCCT GAGCACACTC TTCCTGCAGG CGGCCGAGAT
4651  CTTCGTGGAG TCAGAACTGC CTCTGAGCTG GGCAGACCGG CTGAGTGGCT
4701  GCCTGCGGGG GCCCTGGGCC TGGCTGGTGG TGCTGCTGGC CATGCTGGTG
4751  GAGGTCGCAC TGTGCACCTG GTACCTGGTG GCCTTCCCGC CGGAGGTGGT
4801  GACGGACTGG CACATGCTGC CCACGGAGGC GCTGGTGCAC TGCCGCACAC
4851  GCTCCTGGGT CAGCTTCGGC CTAGCGCACG CCACCAATGC CACGCTGGCC
4901  TTTCTCTGCT TCCTGGGCAC TTTCCTGGTG CGGAGCCAGC CGGGCCGCTA
4951  CAACCGTGCC CGTGGCCTCA CCTTTGCCAT GCTGGCCTAC TTCATCACCT
5001  GGGTCTCCTT TGTGCCCCTC CTGGCCAATG TGCAGGTGGT CCTCAGGCCC
5051  GCCGTGCAGA TGGGCGCCCT CCTGCTCTGT GTCCTGGGCA TCCTGGCTGC
5101  CTTCCACCTG CCCAGGTGTT ACCTGCTCAT GCGGCAGCCA GGGCTCAACA
5151  CCCCCGAGTT CTTCCTGGGA GGGGGCCCTG GGGATGCCCA AGGCCAGAAT
5201  GACGGGAACA CAGGAAATCA GGGGAAACAT GAGTGACCCA ACCCTGTGAT
5251  CTCAGCCCCG GTGAACCCAG ACTTAGCTGC GATCCCCCCC AAGCCAGCAA
5301  TGACCCGTGT CTCGCTACAG AGACCCTCCC GCTCTAGGTT CTGACCCCAG
5351  GTTGTCTCCT GACCCTGACC CCACAGTAAG CCCTAGGCCT GGAGCACGTG
5401  GACACCCCTG TGACCATCTG GGCCCCAGAG CCAAGCTGTG TCCCTGTCCC
5451  TCTGTGCCCA GACCAGGCCT GCCCAGGTAA CCCAGACCCA CTGTTCTGGA
5501  AAGAGGCCCG GAGGGCTCCC AGGGTACCCG CAACCCACAC CGTGAGCTCA
5551  GGAAAAGGAC GCAGGGAGGC CCCGGCCAGA TGGCTGGAAG CCCAAATCAG
5601  GCCCTGCCGA CCTGACCATG TCCCACCAGG GCCCCCATCC TGCACCCTGC
5651  CAGGCACCAC AGCAGTGGGA GGCCAGGTGG GGGCACACAG GCATATGCCC
5701  AGGGCAGAGC CCGCCGAGGT GGGGGTGGCA CCCAGCTTCC TACTCTGCCC
5751  CTTGCCCAGT GGGTAGACAG CATCATGACT GTCACCAGTA CCAGGGACAG
5801  AGCCCAGGTG GGGTGGGGGC GGGGTCCAGC ACCACGGCCA GCACCGACCA
5851  CCAGGACCCC GGAGCCAGCA CCATGGACAG AAAACTGCCC ACCAGGATCT
5901  GACGCCAGCA CGCCGCCAGG CCCACACAGG GTCTCCGGTC AGAGTCCCAG
5951  GGTCAGCTCC CAGCAGGGCC TAGGGGAGGC TGGACCAGCT CCCTGTGCCT
6001  CATTCCAAGG CAGCCCAGCC GGAGAGAAGG GGCACAGGCC ACACATCTGT
6051  CCCATAAAAT TAAACGCTTT TTAGTGTTTA AAATAAGCAG CATTTACACA
6101  GAAGCAGCTC TATGTTAACC ATCTAAACGC TGGGACTTTG ATACAGTATC
6151  TACAGCACAG ACACGTGGGG GCCAGAGAAG CCAGGAAGGC CGCGATGTGT
6201  GCGCGCAGTG TGTGCACTCA CCAAGGACGG GCCACCTGAC TGCCCATCTC
6251  CCCAAGACCT CCCTCCCTGT GGCAGCTGTG CACATCGGGG CCCCTTGACC
6301  CTGCGGGCCA TGGTCCCTCC CTGCCCTGGC TGGGACACGG TGGGCAGGAT
6351  GTCCAGCCCT CTCTCGTCTT CCGGTCCCGT CCTCCACGTA CTTTCAGACT
6401  GTTGCCGGAT GGGAGGAGAG AAGGTGCAGG CTGCTCCAAG GGGCAGCAGC
6451  AGGTGGGACA GATGACAGGG TCGCCTCCTC CCCCGAGTCA CCGGCCAGGC
6501  AATAAATAAA TAAATTAGAT CCCTACTCCA GACAGGGGGC CTGTGCACCG
6551  CAGGGGGTTG CCCCGCGTGG ACCACCCTGG GGGCCTGGGC ACAGCTGTGC
6601  AGGAGGTGGG GGTCAGCCGA GAGCCCGAGG GGGTCTTCCT CATCCCAGGA
6651  GGGATCCCCA CCAGACACAA GGGGTTGGGA GGTCCGAGGC TCTCGCTGAG
```

```
    6701  GGGCAGAGAG GGAGCGCCCC CAACACGGCT GCTCAGACAC AGGTGCTGTC
    6751  AGGAGCTGGA GCAGCCAGGC TGCCAGGGCA GATGGTGGTG GTCCAGCCTG
    6801  CCCCCCACCC TGCCTCCCGT CCTGGCCCCC ACGAAGGCAA GCCCACGCGA
    6851  GCTCTGCATG CGGCAGGACC GCCAGCTCCC CACCTCAGGC AGGGCTGGGG
    6901  CATGCGCCAC GAGTCACATG ATGTCCACGA AGAACTCGCA GGGGTTCCCC
    6951  ATAGCCTTCT GGAAGGACTG GCGGCTGCCT GTCAATTCCG GGGGGACGGC
    7001  AGCCAGCTCC CGGACAGGGG GTCCCCCGGG TGGCCCCCCC ACCACTGTAT
    7051  AGGCCTTGGT CGTGGGGTGG GGCGGGGGGA GCCCCGGGGC GGTAGCCGAG
    7101  GCCTGACTGC GTGGGCTGCT GCCACGGCTG AGCTGGCCGG CCGGACGCTC
    7151  TCGCCAGCTG CTCCCCACCC CACTCGGTGC CGTGTGATCC GATTCACTGC
    7201  CACTGCCCCC AGCTCCCGCC GCCCGACGCT CCTTCTCACG GCCCGGGGCC
    7251  CGGCGGCTGC TCCGGGTGGA CCCACTGCTT TTGCTCCCTG GGAGTGAGAA
    7301  CAGGATGGGG AAGGAGCCTG TCAGCACCAG GCCCGGCCAC GTCCAGAACA
    7351  ACCGCCCCCG CCAGGTGACC GGGCAGGAAC GGTGGCCCGT GCAGGACGGG
    7401  TGGGTGGGGC AGCCATGCAG GCGCGAGGAC AGGGCCGGCA CCCCCAGGGT
    7451  CGCAGGGGCA GCTGCGGCAG GCATGGGCTG GTGGGGGCAG TGGGAGGCAG
    7501  GCAGGGTCCT GGGGCAAGCT GGGCGCCCCC ACACCTCACC CCGATGCTTG
    7551  AGCTGAGGGC CTGGTCTTGC GGCTGAAAGA CTGAGGTGCC AGCGAGGGCC
    7601  CCTCATGCCC GGTGCCCCCG TGGCCATCCT GGATTCCCCA CCCAAGGCCC
    7651  CACTGTCCCC CCGGCCCAGG ACCCTGGCCG ACGGATGACT CAGCTCAGCC
    7701  CTGTCCTGGG CTCCCAAGAC GCAGTGGGAG CTGGAGGGCG TGGCTGGCTG
    7751  GGGACATGCT GAGGGACCCC GGGCGGGACC CTGGCTTACC GGCCCAAGGT
    7801  CCGCTCCGCT AGTCCTTCAG TCTAAGGCTT GTTTAGCACA AGACAAGGGA
    7851  TAGCACGAGT TACACGCCCG GCTGCCTGGC ACCTGCCCGG CACCCACCCG
    7901  CCACCAGTGG GGACTGACCG CGGGCTGGGC GGGGCTGAAG TGGGCGCAAG
    7951  CGCCGGTCGA GGGTCAGTAG ACACCCAGCG TGGCTTCTCT GCGGTCCCAC
    8001  A
(SEQ ID NO:3)
```

FEATURES:

Start:	2118
Exon:	2118-2308
Intron:	2309-2409
Exon:	2410-2710
Intron:	2711-2795
Exon:	2796-3578
Intron:	3579-3692
Exon:	3693-3896
Intron:	3897-4030
Exon:	4031-4151
Intron:	4152-4277
Exon:	4278-5236
Stop:	5234

MAP POSITION:

#	SHGCNAME	CHROM#	LOD_SCORE	DIST.(cRs)
1	SHGC-57364	1	6.42	32

Bac Accession #: AC026283

ALLELIC VARIANTS (SNPs):

Position	Major	Minor	Context
406 (SEQ ID NO:5)	t	c	gccatctgtgcaatgtgcaggtgccagttgcatgtgccatgcgtgttggc[t/c]gtgagcgtgtgctgttttcgtgtatgtgccatgcacgtatgtgctgcgtg
852 (SEQ ID NO:6)	t	c	attcgtgtgcaccgtacacgtgtgccatttgcatgtatgctgtgcacgtg[t/c]ggcatgcatgtgtgccgtttgcatgccatgcatgtgttccttgcgtgtgc
897 (SEQ ID NO:7)	g	a	acgtgcggcatgcatgtgtgccgtttgcatgccatgcatgtgttccttgc[g/a]tgtgccgtgcgtgtcccatgcacgtgtgccgtgcatgtgccattcgcgtg
1,433 (SEQ ID NO:8)	c	t	cgcaggagtggcccagacgggggtcagcagtgtgagtacagctggccatg[c/t]ggttcctacagcttccaggcgtcagactctggcagaagggctgagaccct
5,845 (SEQ ID NO:9)	t	c	ggacagagcccaggtggggtgggggcggggtccagcaccacggccagcac[t/c]gaccaccaggaccccggagccagcaccatggacagaaaactgcccaccag
7,028 (SEQ ID NO:10)	g	a	cctgtcaattccgggggggacggcagccagctcccggacaggggtccccc[g/a]ggtggcccccccaccactgtataggccttggtcgtggggtggggcggggg

Position	Allele 1	Allele 2	
406	t	c	Intron
852	t	c	Intron
897	g	a	Intron
1,433	c	t	Intron
5,845	t	c	Intron
7,028	g	a	Intron

SEQUENCE LISTING

<110> Ming-Hui WEI, Wenyan ZHONG, Karen A. KETCHUM, Valentina DIFRANCESCO, Ellen M. BEASLEY

<120> ISOLATED HUMAN G-PROTEIN COUPLED RECEPTORS, NUCLEIC ACID MOLECULES ENCODING HUMAN GPCR PROTEINS, AND USES THEREOF

<130> CL000869PCT

<160> 10

<170> FastSEQ for Windows Version 4.0

<210> 1
<211> 3458
<212> DNA
<213> HUMAN

<400> 1
cccgggctgc aggaattcgg cacgaggcca tgctgggccc tgctgtcctg ggcctcagcc 60
tctgggctct cctgcaccct gggacggggg ccccattgtg cctgtcacag caacttagga 120
tgaaggggga ctacgtgctg gggggggctgt tccccctggg cgaggccgag gaggctggcc 180
tccgcagccg gacacggccc agcagccctg tgtgcaccag gttctcctca aacggcctgc 240
tctgggcact ggccatgaaa atggccgtgg aggagatcaa caacaagtcg gatctgctgc 300
ccgggctgcg cctgggctac gacctctttg atacgtgctc ggagcctgtg gtggccatga 360
agcccagcct catgttcctg gccaaggcag gcagccgcga catcgccgcc tactgcaact 420
acacgcagta ccagccccgt gtgctggctg tcatcgggcc ccactcgtca gagctcgcca 480
tggtcaccgg caagttcttc agcttcttcc tcatgcccca ggtcagctac ggtgctagca 540
tggagctgct gagcgcccgg gagaccttcc cctccttctt ccgcaccgtg cccagcgacc 600
gtgtgcagct gacggccgcc gcggagctgc tgcaggagtt cggctggaac tgggtggccg 660
ccctgggcag cgacgacgag tacggccggc agggcctgag catcttctcg gccctggccg 720
cggcacgcgg catctgcatc gcgcacgagg gcctggtgcc gctgccccgt gccgatgact 780
cgcggctggg gaaggtgcag gacgtcctgc accaggtgaa ccagagcagc gtgcaggtgg 840
tgctgctgtt cgcctccgtg cacgccgccc acgccctctt caactacagc atcagcagca 900
ggctctcgcc caaggtgtgg gtggccagcg aggcctggct gacctctgac ctggtcatgg 960
ggctgcccgg catggcccag atgggcacgg tgcttggctt cctccagagg ggtgcccagc 1020
tgcacgagtt cccccagtac gtgaagacgc acctggccct ggccaccgac ccggccttct 1080
gctctgccct gggcgagagg gagcagggtc tggaggagga cgtggtgggc cagcgctgcc 1140
cgcagtgtga ctgcatcacg ctgcagaacg tgagcgcagg gctaaatcac caccagacgt 1200
tctctgtcta cgcagctgtg tatagcgtgg cccaggccct gcacaacact cttcagtgca 1260
acgcctcagg ctgccccgcg caggaccccg tgaagccctg gcagctcctg gagaacatgt 1320
acaacctgac cttccacgtg ggcgggctgc cgctgcggtt cgacagcagc ggaaacgtgg 1380
acatggagta cgacctgaag ctgtgggtgt ggcagggctc agtgcccagg ctccacgacg 1440
tgggcaggtt caacggcagc ctcaggacag agcgcctgaa gatccgctgg cacacgtctg 1500
acaaccagaa gcccgtgtcc cggtgctcgc ggcagtgcca ggagggccag gtgcgccggg 1560
tcaaggggtt ccactcctgc tgctacgact gtgtggactg cgaggcgggc agctaccggc 1620
aaaacccaga cgacatcgcc tgcacctttt gtggccagga tgagtggtcc ccggagcgaa 1680
gcacacgctg cttccgccgc aggtctcggt tcctggcatg gggcgagccg gctgtgctgc 1740
tgctgctcct gctgctgagc ctggcgctgg gccttgtgct ggctgctttg gggctgttcg 1800
ttcaccatcg ggacagccca ctggttcagg cctcgggggg gcccctggcc tgctttggcc 1860
tggtgtgcct gggcctggtc tgcctcagcg tcctcctgtt ccctggccag cccagccctg 1920
cccgatgcct ggcccagcag cccttgtccc acctcccgct cacgggctgc ctgagcacac 1980
tcttcctgca ggcggccgag atcttcgtgg agtcagaact gcctctgagc tgggcagacc 2040
ggctgagtgg ctgcctgcgg gggccctggg cctggctggt ggtgctgctg gccatgctgg 2100
tggaggtcgc actgtgcacc tggtacctgg tggccttccc gccggaggtg gtgacggact 2160
ggcacatgct gcccacggag gcgctggtgc actgccgcac acgctcctgg gtcagcttcg 2220
gcctagcgca cgccaccaat gccacgctgg cctttctctg cttcctgggc actttcctgg 2280
tgcggagcca gccgggccgc tacaaccgtg cccgtggcct caccctttgcc atgctggcct 2340
acttcatcac ctgggtctcc tttgtgcccc tcctggccaa tgtgcaggtg gtcctcaggc 2400

```
ccgccgtgca gatgggcgcc ctcctgctct gtgtcctggg catcctggct gccttccacc 2460
tgcccaggtg ttacctgctc atgcggcagc cagggctcaa caccccgag ttcttcctgg 2520
gaggggccc tggggatgcc caaggccaga atgacgggaa cacaggaaat caggggaaac 2580
atgagtgacc caaccctgtg atctcagccc cggtgaaccc agacttagct gcgatccccc 2640
ccaagccagc aatgacccgt gtctcgctac agagaccctc ccgctctagg ttctgacccc 2700
aggttgtctc ctgaccctga ccccacagtg agccctaggc ctggagcacg tggacacccc 2760
tgtgaccatc tgggccccag agccaagctg tgtccctgtc cctctgtgcc cagaccaggc 2820
ctgcccaggt aacccagacc cactgttctg gaaagaggcc cggagggctc ccagggtacc 2880
cgcaacccac accgtgagct caggaaaagg acgcagggag gccccggcca gatggctgga 2940
agcccaaatc aggccctgcc gacctgacca tgtcccacca gggcccccat cctgcaccct 3000
gccaggcacc acagcagtgg gaggccaggt gggggcacac aggcatatgc ccagggcaga 3060
gcccgccgag gtgggggtgg cacccagctt cctactctgc cccttgccca gtgggtagac 3120
agcatcatga ctgtcaccag taccagggac agagcccagg tggggtgggg gcggggtcca 3180
gcaccacggc cagcaccgac caccaggacc ccggagccag caccatggac agaaaactgc 3240
ccaccaggat ctgacgccag cacgccgcca ggcccacaca gggtctccgg tcagagtccc 3300
agggtcagct cccagcaggg cctaggggag gctggaccag ctccctgtgc ctcattccaa 3360
ggcagcccag ccggagagaa ggggcacagg ccacacatct gtcccataaa attaaacgct 3420
ttttagtgtt taaaataaaa aaaaaaaaaa aaaaaaaa                         3458
```

<210> 2
<211> 852
<212> PRT
<213> HUMAN

<400> 2

```
Met Leu Gly Pro Ala Val Leu Gly Leu Ser Leu Trp Ala Leu Leu His
 1               5                  10                  15
Pro Gly Thr Gly Ala Pro Leu Cys Leu Ser Gln Gln Leu Arg Met Lys
            20                  25                  30
Gly Asp Tyr Val Leu Gly Gly Leu Phe Pro Leu Gly Glu Ala Glu Glu
        35                  40                  45
Ala Gly Leu Arg Ser Arg Thr Arg Pro Ser Ser Pro Val Cys Thr Arg
    50                  55                  60
Phe Ser Ser Asn Gly Leu Leu Trp Ala Leu Ala Met Lys Met Ala Val
65                  70                  75                  80
Glu Glu Ile Asn Asn Lys Ser Asp Leu Leu Pro Gly Leu Arg Leu Gly
                85                  90                  95
Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val Ala Met Lys Pro
            100                 105                 110
Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp Ile Ala Ala Tyr
        115                 120                 125
Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala Val Ile Gly Pro
    130                 135                 140
His Ser Ser Glu Leu Ala Met Val Thr Gly Lys Phe Phe Ser Phe Phe
145                 150                 155                 160
Leu Met Pro Gln Val Ser Tyr Gly Ala Ser Met Glu Leu Leu Ser Ala
                165                 170                 175
Arg Glu Thr Phe Pro Ser Phe Phe Arg Thr Val Pro Ser Asp Arg Val
            180                 185                 190
Gln Leu Thr Ala Ala Ala Glu Leu Leu Gln Glu Phe Gly Trp Asn Trp
        195                 200                 205
Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr Gly Arg Gln Gly Leu Ser
    210                 215                 220
Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly Ile Cys Ile Ala His Glu
225                 230                 235                 240
Gly Leu Val Pro Leu Pro Arg Ala Asp Asp Ser Arg Leu Gly Lys Val
                245                 250                 255
Gln Asp Val Leu His Gln Val Asn Gln Ser Ser Val Gln Val Val Leu
            260                 265                 270
Leu Phe Ala Ser Val His Ala Ala His Ala Leu Phe Asn Tyr Ser Ile
        275                 280                 285
```

```
Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala Ser Glu Ala Trp Leu
    290                 295                 300
Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met Ala Gln Met Gly Thr
305                 310                 315                 320
Val Leu Gly Phe Leu Gln Arg Gly Ala Gln Leu His Glu Phe Pro Gln
                325                 330                 335
Tyr Val Lys Thr His Leu Ala Leu Ala Thr Asp Pro Ala Phe Cys Ser
            340                 345                 350
Ala Leu Gly Glu Arg Glu Gln Gly Leu Glu Glu Asp Val Val Gly Gln
        355                 360                 365
Arg Cys Pro Gln Cys Asp Cys Ile Thr Leu Gln Asn Val Ser Ala Gly
    370                 375                 380
Leu Asn His His Gln Thr Phe Ser Val Tyr Ala Ala Val Tyr Ser Val
385                 390                 395                 400
Ala Gln Ala Leu His Asn Thr Leu Gln Cys Asn Ala Ser Gly Cys Pro
                405                 410                 415
Ala Gln Asp Pro Val Lys Pro Trp Gln Leu Leu Glu Asn Met Tyr Asn
            420                 425                 430
Leu Thr Phe His Val Gly Gly Leu Pro Leu Arg Phe Asp Ser Ser Gly
        435                 440                 445
Asn Val Asp Met Glu Tyr Asp Leu Lys Leu Trp Val Trp Gln Gly Ser
    450                 455                 460
Val Pro Arg Leu His Asp Val Gly Arg Phe Asn Gly Ser Leu Arg Thr
465                 470                 475                 480
Glu Arg Leu Lys Ile Arg Trp His Thr Ser Asp Asn Gln Lys Pro Val
                485                 490                 495
Ser Arg Cys Ser Arg Gln Cys Gln Glu Gly Gln Val Arg Arg Val Lys
            500                 505                 510
Gly Phe His Ser Cys Cys Tyr Asp Cys Val Asp Cys Glu Ala Gly Ser
        515                 520                 525
Tyr Arg Gln Asn Pro Asp Asp Ile Ala Cys Thr Phe Cys Gly Gln Asp
    530                 535                 540
Glu Trp Ser Pro Glu Arg Ser Thr Arg Cys Phe Arg Arg Arg Ser Arg
545                 550                 555                 560
Phe Leu Ala Trp Gly Glu Pro Ala Val Leu Leu Leu Leu Leu Leu Leu
                565                 570                 575
Ser Leu Ala Leu Gly Leu Val Leu Ala Ala Leu Gly Leu Phe Val His
            580                 585                 590
His Arg Asp Ser Pro Leu Val Gln Ala Ser Gly Gly Pro Leu Ala Cys
        595                 600                 605
Phe Gly Leu Val Cys Leu Gly Leu Val Cys Leu Ser Val Leu Leu Phe
    610                 615                 620
Pro Gly Gln Pro Ser Pro Ala Arg Cys Leu Ala Gln Gln Pro Leu Ser
625                 630                 635                 640
His Leu Pro Leu Thr Gly Cys Leu Ser Thr Leu Phe Leu Gln Ala Ala
                645                 650                 655
Glu Ile Phe Val Glu Ser Glu Leu Pro Leu Ser Trp Ala Asp Arg Leu
            660                 665                 670
Ser Gly Cys Leu Arg Gly Pro Trp Ala Trp Leu Val Val Leu Leu Ala
        675                 680                 685
Met Leu Val Glu Val Ala Leu Cys Thr Trp Tyr Leu Val Ala Phe Pro
    690                 695                 700
Pro Glu Val Val Thr Asp Trp His Met Leu Pro Thr Glu Ala Leu Val
705                 710                 715                 720
His Cys Arg Thr Arg Ser Trp Val Ser Phe Gly Leu Ala His Ala Thr
                725                 730                 735
Asn Ala Thr Leu Ala Phe Leu Cys Phe Leu Gly Thr Phe Leu Val Arg
            740                 745                 750
Ser Gln Pro Gly Arg Tyr Asn Arg Ala Arg Gly Leu Thr Phe Ala Met
        755                 760                 765
```

```
Leu Ala Tyr Phe Ile Thr Trp Val Ser Phe Val Pro Leu Leu Ala Asn
    770                 775                 780
Val Gln Val Val Leu Arg Pro Ala Val Gln Met Gly Ala Leu Leu Leu
785                 790                 795                 800
Cys Val Leu Gly Ile Leu Ala Ala Phe His Leu Pro Arg Cys Tyr Leu
                805                 810                 815
Leu Met Arg Gln Pro Gly Leu Asn Thr Pro Glu Phe Phe Leu Gly Gly
            820                 825                 830
Gly Pro Gly Asp Ala Gln Gly Gln Asn Asp Gly Asn Thr Gly Asn Gln
        835                 840                 845
Gly Lys His Glu
    850
```

<210> 3
<211> 8001
<212> DNA
<213> Human

<400> 3

```
gcaaaagagg agaggcctgc gtgggatgtg tgcgtggagg tgggggccac tcccagccac 60
aacagtggcc tggacagaaa ggggatgcaa ttcagcagag tcttcctgga tgtcaccccc 120
acctcagggt ctgtgggagc tgcatatggt gccggcaaag gctgccgact gcagtatggg 180
ccgggagaac tgcctgggtc tgcgtgggcc ccagggcagg gctccctccg ggtgttgcct 240
tctgtacaag tgccatgctt gtgccgtttg cgtgtcccaa gtgcgagtgt gctatttgcg 300
tgtgccgcac gtgtgccgtt tgcatgtgct gtttgcatgt accatgtgca tgtgtgccat 360
ctgtgcaatg tgcaggtgcc agttgcatgt gccatgcgtg ttggctgtga gcgtgtgctg 420
ttttcgtgta tgtgccatgc acgtatgtgc tgcgtgttgg ccgtgcacgt gtgccacgtg 480
catgtgtgcc atttgtgtat gccgtgtgtg ctgtgagcgt atgctgtgcg catgtgcgtg 540
ccatgcgctt tgcgtgtacc atgtgtgtgc tgtttgcatg tgccatttgt gtcatgcacg 600
tgccatttgc gtgcgatgcg cgtgtgccac acgtcgtttg cttgcgtgcc atgcatgtgt 660
ggcatttgtg tatgtgccgt ttccgtgtgt attgtgtgtg ccgtgtgtgt gccatttgca 720
tgtgccgttt tgtgtgtgcc atgcgcgtgt gccatgcact tgccgtgcgt gtgccatttg 780
tgtgtaccat gcgcatgtgc cattcgtgtg caccgtacac gtgtgccatt tgcatgtatg 840
ctgtgcacgt gcggcatgca tgtgtgccgt ttgcatgcca tgcatgtgtt ccttgcgtgt 900
gccgtgcgtg tcccatgcac gtgtgccgtg catgtgccat tcgcgtgtac catgcgcatg 960
tgccgtttgt gtgtgtgccg cattcctgtg ctcgtgtgcc aggttcgcat gtgcgccata 1020
ttcacgtgtg ctcagcatgt gccatatgca tgtgcggtgg tagtgtgtgt ccctcacagg 1080
tcctcctcac aacaccatgg ggaagaagca ccagccaggg cacacctcct ggtatctgct 1140
aggtctgcca ggccctagct gaagctgagt gccccccagt tcccctggga gggcctgcgc 1200
ctggagtctg ctgtgtcccc gagggcaccc ccaaagcaac acagaggcag aggagtcccg 1260
gccctgcaca cctggtgctg ctccagctgc cgctcatttg cctgtggccc ttcctccctt 1320
gtttgcgtgc ccccctggca aacaaactct acccccagca ggagccacct gtgtgcctgc 1380
cacgcaggag tggcccagac gggggtcagc agtgtgagta cagctggcca tgcggttcct 1440
acagcttcca ggcgtcagac tctggcagaa gggctgagac cctcaaggaa ctctgctccc 1500
aagcagactg ggagggcagc accaccaccc cagggccctc cccagctgca gggtggaggc 1560
ctggctggcc ggctgcccac tggcctgact ggtctgcagg cctagggggc ccatccctgc 1620
tgcccccggc tccggccagc acagccttga gtgggagcca gaagctcccg gggctgggta 1680
ggaggcattt ctgtgcttat gaaaagcccc agggctgggt gtctctgcat ccctcccacg 1740
cagctgagac ctcagagccc tggaggcccc tttgcccccc ctcctctcca cagcctgctg 1800
ggcaactcca ggaatcgggg ggtggcaagg ggctcagcca caggcaggga acaaggccac 1860
ggccagcgac tgagcagagc ctgcctgccg gtcaacgctg gccatagagc ctggcagtgg 1920
cctcaggcag agtctgacgc gcacaaactt tcaggcccag gaagcgagga caccactggg 1980
gccccagggt gtggcaagtg aggatggcaa gggttttgct aaacaaatcc tctgcccgct 2040
ccccgccccg ggctcactcc atgtgaggcc ccagtcgggg cagccacctg ccgtgcctgt 2100
tggaagttgc ctctgccatg ctgggccctg ctgtcctggg cctcagcctc tgggctctcc 2160
tgcaccctgg gacgggggcc ccattgtgcc tgtcacagca acttaggatg aaggggggact 2220
acgtgctggg ggggctgttc cccctgggcg aggccgagga ggctggcctc cgcagccgga 2280
cacggcccag cagccctgtg tgcaccaggt acagaggtgg gacggcctgg gtcggggtca 2340
gggtgaccag gtctggggtg ctcctgagct ggggccgagg tggccatctg cggttctgtg 2400
tggccccagg ttctcctcaa acggcctgct ctgggcactg gccatgaaaa tggccgtgga 2460
```

ggagatcaac aacaagtcgg atctgctgcc cgggctgcgc ctgggctacg acctctttga 2520
tacgtgctcg gagcctgtgg tggccatgaa gcccagcctc atgttcctgg ccaaggcagg 2580
cagccgcgac atcgccgcct actgcaacta cacgcagtac cagccccgtg tgctggctgt 2640
catcgggccc cactcgtcag agctcgccat ggtcaccggc aagttcttca gcttcttcct 2700
catgccccag gtgggcgccc cccaccatca cccaccccca cccagccctg cccgtgggag 2760
cccctgtgtc aggagatgcc tcttggccct tgcaggtcag ctacggtgct agcatggagc 2820
tgctgagcgc ccgggagacc ttcccctcct tcttccgcac cgtgcccagc gaccgtgtgc 2880
agctgacggc cgccgcggag ctgctgcagg agttcggctg gaactgggtg gccgccctgg 2940
gcagcgacga cgagtacggc cggcagggcc tgagcatctt ctcggccctg gccgcggcac 3000
gcggcatctg catcgcgcac gagggcctgg tgccgctgcc ccgtgccgat gactcgcggc 3060
tggggaaggt gcaggacgtc ctgcaccagg tgaaccagag cagcgtgcag gtggtgctgc 3120
tgttcgcctc cgtgcacgcc gcccacgccc tcttcaacta cagcatcagc agcaggctct 3180
cgcccaaggt gtgggtggcc agcgaggcct ggctgacctc tgacctggtc atggggctgc 3240
ccggcatggc ccagatgggc acggtgcttg gcttcctcca gaggggtgcc cagctgcacg 3300
agttccccca gtacgtgaag acgcacctgg ccctggccac cgacccggcc ttctgctctg 3360
ccctgggcga gagggagcag ggtctggagg aggacgtggt gggccagcgc tgcccgcagt 3420
gtgactgcat cacgctgcag aacgtgagcg cagggctaaa tcaccaccag acgttctctg 3480
tctacgcagc tgtgtatagc gtggcccagg ccctgcacaa cactcttcag tgcaacgcct 3540
caggctgccc cgcgcaggac cccgtgaagc cctggcaggt gagcccggga gatgggggtg 3600
tgctgtcctc tgcatgtgcc caggccacca ggcacggcca ccacgcctga gctggaggtg 3660
gctggcggct cagccccgtc ccccgcccgc agctcctgga gaacatgtac aacctgacct 3720
tccacgtggg cgggctgccg ctgcggttcg acagcagcgg aaacgtggac atggagtacg 3780
acctgaagct gtgggtgtgg cagggctcag tgcccaggct ccacgacgtg ggcaggttca 3840
acggcagcct caggacagag cgcctgaaga tccgctggca cacgtctgac aaccaggtga 3900
ggtgagggtg ggtgtgccag gcgtgcccgt ggtagccccc gcggcagggc gcagcctggg 3960
ggtgggggcc gttccagtct cccgtgggca tgcccagccg agcagagcca gaccccaggc 4020
ctgtgcgcag aagcccgtgt cccggtgctc gcggcagtgc caggagggcc aggtgcgccg 4080
ggtcaagggg ttccactcct gctgctacga ctgtgtggac tgcgaggcgg gcagctaccg 4140
gcaaaaccca ggtgagccgc cttcccggca ggcgggggtg ggaacgcagc aggggagggt 4200
cctgccaagt cctgactctg agaccagagc ccacagggga caagacgaac acccagcgcc 4260
cttctcctct ctcacagacg acatcgcctg caccttttgt ggccaggatg agtggtcccc 4320
ggagcgaagc acacgctgct tccgccgcag gtctcggttc ctggcatggg gcgagccggc 4380
tgtgctgctg ctgctcctgc tgctgagcct ggcgctgggc cttgtgctgg ctgctttggg 4440
gctgttcgtt caccatcggg acagcccact ggttcaggcc tcgggggggc ccctggcctg 4500
ctttggcctg gtgtgcctgg gcctggtctg cctcagcgtc ctcctgttcc ctggccagcc 4560
cagccctgcc cgatgcctgg cccagcagcc cttgtcccac ctcccgctca cgggctgcct 4620
gagcacactc ttcctgcagg cggccgagat cttcgtggag tcagaactgc ctctgagctg 4680
ggcagaccgg ctgagtggct gcctgcgggg gccctgggcc tggctggtgg tgctgctggc 4740
catgctggtg gaggtcgcac tgtgcacctg gtacctggtg gccttcccgc cggaggtggt 4800
gacggactgg cacatgctgc ccacggaggc gctggtgcac tgccgcacac gctcctgggt 4860
cagcttcggc ctagcgcacg ccaccaatgc cacgctggcc tttctctgct tcctgggcac 4920
tttcctggtg cggagccagc cgggccgcta caaccgtgcc cgtggcctca cctttgccat 4980
gctggcctac ttcatcacct gggtctcctt tgtgcccctc ctggccaatg tgcaggtggt 5040
cctcaggccc gccgtgcaga tgggcgccct cctgctctgt gtcctgggca tcctggctgc 5100
cttccacctg cccaggtgtt acctgctcat gcggcagcca gggctcaaca cccccgagtt 5160
cttcctggga gggggccctg gggatgccca aggccagaat gacgggaaca caggaaatca 5220
ggggaaacat gagtgaccca accctgtgat ctcagccccg gtgaacccag acttagctgc 5280
gatccccccc aagccagcaa tgacccgtgt ctcgctacag agaccctccc gctctaggtt 5340
ctgaccccag gttgtctcct gaccctgacc ccacagtaag ccctaggcct ggagcacgtg 5400
gacacccctg tgaccatctg ggccccagag ccaagctgtg tccctgtccc tctgtgccca 5460
gaccaggcct gcccaggtaa cccagaccca ctgttctgga aagaggcccg gagggctccc 5520
agggtacccg caacccacac cgtgagctca ggaaaaggac gcagggaggc cccggccaga 5580
tggctggaag cccaaatcag gccctgccga cctgaccatg tcccaccagg gcccccatcc 5640
tgcaccctgc caggcaccac agcagtggga ggccaggtgg gggcacacag gcatatgccc 5700
agggcagagc ccgccgaggt gggggtggca cccagcttcc tactctgccc cttgcccagt 5760
gggtagacag catcatgact gtcaccagta ccagggacag agcccaggtg gggtgggggc 5820
ggggtccagc accacggcca gcaccgacca ccaggacccc ggagccagca ccatggacag 5880
aaaactgccc accaggatct gacgccagca cgccgccagg cccacacagg gtctccggtc 5940
agagtcccag ggtcagctcc cagcagggcc taggggaggc tggaccagct ccctgtgcct 6000
cattccaagg cagcccagcc ggagagaagg ggcacaggcc acacatctgt cccataaaat 6060
taaacgcttt ttagtgttta aaataagcag catttacaca gaagcagctc tatgttaacc 6120

```
atctaaacgc tgggactttg atacagtatc tacagcacag acacgtgggg gccagagaag 6180
ccaggaaggc cgcgatgtgt gcgcgcagtg tgtgcactca ccaaggacgg gccacctgac 6240
tgcccatctc cccaagacct ccctccctgt ggcagctgtg cacatcgggg ccccttgacc 6300
ctgcgggcca tggtccctcc ctgccctggc tgggacacgg tgggcaggat gtccagccct 6360
ctctcgtctt ccggtcccgt cctccacgta ctttcagact gttgccggat gggaggagag 6420
aaggtgcagg ctgctccaag gggcagcagc aggtgggaca gatgacaggg tcgcctcctc 6480
ccccgagtca ccggccaggc aataaataaa taaattagat ccctactcca gacaggggggc 6540
ctgtgcaccg caggggttg ccccgcgtgg accaccctgg gggcctgggc acagctgtgc 6600
aggaggtggg ggtcagccga gagcccgagg gggtcttcct catcccagga gggatcccca 6660
ccagacacaa ggggttggga ggtccgaggc tctcgctgag gggcagagag ggagcgcccc 6720
caacacggct gctcagacac aggtgctgtc aggagctgga gcagccaggc tgccagggca 6780
gatggtggtg gtccagcctg ccccccaccc tgcctcccgt cctggccccc acgaaggcaa 6840
gcccacgcga gctctgcatg cggcaggacc gccagctccc cacctcaggc agggctgggg 6900
catgcgccac gagtcacatg atgtccacga agaactcgca ggggttcccc atagccttct 6960
ggaaggactg gcggctgcct gtcaattccg gggggacggc agccagctcc cggacagggg 7020
gtcccccggg tggccccccc accactgtat aggccttggt cgtggggtgg ggcgggggga 7080
gccccggggc ggtagccgag gcctgactgc gtgggctgct gccacggctg agctggccgg 7140
ccggacgctc tcgccagctg ctccccaccc cactcggtgc cgtgtgatcc gattcactgc 7200
cactgccccc agctcccgcc gcccgacgct ccttctcacg gcccggggcc cggcggctgc 7260
tccgggtgga cccactgctt ttgctccctg ggagtgagaa caggatgggg aaggagcctg 7320
tcagcaccag gcccggccac gtccagaaca accgccccg ccaggtgacc gggcaggaac 7380
ggtggcccgt gcaggacggg tgggtggggc agccatgcag gcgcgaggac agggccggca 7440
cccccagggt cgcaggggca gctgcggcag gcatgggctg gtgggggcag tgggaggcag 7500
gcagggtcct ggggcaagct gggcgccccc acacctcacc ccgatgcttg agctgagggc 7560
ctggtcttgc ggctgaaaga ctgaggtgcc agcgagggcc cctcatgccc ggtgcccccg 7620
tggccatcct ggattcccca cccaaggccc cactgtcccc ccggcccagg accctggccg 7680
acggatgact cagctcagcc ctgtcctggg ctcccaagac gcagtgggag ctggagggcg 7740
tggctggctg gggacatgct gagggacccc gggcgggacc ctggcttacc ggcccaaggt 7800
ccgctccgct agtccttcag tctaaggctt gtttagcaca agacaaggga tagcacgagt 7860
tacacgcccg gctgcctggc acctgcccgg cacccacccg ccaccagtgg ggactgaccg 7920
cgggctgggc ggggctgaag tgggcgcaag cgccggtcga gggtcagtag acacccagcg 7980
tggcttctct gcggtcccac a                                          8001
```

<210> 4
<211> 1138
<212> PRT
<213> Rattus norvegicus

<400> 4

```
Ser Leu Pro Gly Asp Phe Leu Leu Ala Gly Leu Phe Ser Leu His Gly
 1               5                  10                  15
Asp Cys Leu Gln Val Arg His Arg Pro Leu Val Thr Ser Cys Asp Arg
            20                  25                  30
Pro Asp Ser Phe Asn Gly His Gly Tyr His Leu Phe Gln Ala Met Arg
        35                  40                  45
Phe Thr Val Glu Glu Ile Asn Asn Ser Gln Ser Asp Leu Leu Pro Gly
    50                  55                  60
Leu Arg Leu Gly Tyr Asp Leu Phe Asp Thr Cys Ser Glu Pro Val Val
65                  70                  75                  80
Ala Met Lys Pro Ser Leu Met Phe Leu Ala Lys Ala Gly Ser Arg Asp
                85                  90                  95
Ile Ala Ala Tyr Cys Asn Tyr Thr Gln Tyr Gln Pro Arg Val Leu Ala
            100                 105                 110
Val Ile Gly Pro His Ser Ser Leu Leu Pro Leu Gly Tyr Leu Asp Cys
        115                 120                 125
Ser Glu Leu Leu Ala Gly Arg Ile Val Ala Ile Gly Pro Ser Ser Ala
    130                 135                 140
Leu Leu Pro Asn Ile Thr Leu Gly Tyr Glu Leu Tyr Asp Val Cys Ser
145                 150                 155                 160
Glu Ser Ala Asn Val Tyr Ala Thr Leu Arg Val Leu Ala Leu Gln Gly
                165                 170                 175
```

```
Pro Arg His Ile Glu Ile Gln Lys Asp Leu Arg Asn His Ser Ser Lys
                180                 185                 190
Val Val Ala Phe Ile Gly Pro Asp Asn Gln Ser Glu Leu Ala Met Val
            195                 200                 205
Thr Gly Lys Phe Phe Ser Phe Phe Leu Met Pro Gln Val Ser Tyr Gly
    210                 215                 220
Ala Ser Met Glu Leu Leu Ser Ala Arg Glu Thr Phe Pro Ser Phe Phe
225                 230                 235                 240
Arg Thr Val Pro Ser Asp Arg Val Gln Leu Thr Ala Ala Ala Glu Leu
                245                 250                 255
Leu Gln Glu Phe Gly Trp Ala Thr Phe Leu Met Pro Val Ser Tyr Ala
            260                 265                 270
Ser Leu Ser Ala Phe Pro Ser Phe Arg Thr Val Pro Ser Asp Arg Gln
        275                 280                 285
Leu Leu Gln Phe Gly Trp Ser Thr Asp His Ala Val Thr Thr Ala Ala
    290                 295                 300
Leu Leu Gly Pro Phe Leu Met Pro Leu Val Ser Tyr Glu Ala Ser Ser
305                 310                 315                 320
Val Val Leu Ser Ala Lys Arg Lys Phe Pro Ser Phe Leu Arg Thr Val
                325                 330                 335
Pro Ser Asp Arg His Gln Val Glu Val Met Val Gln Leu Leu Gln Ser
            340                 345                 350
Phe Gly Trp Gln Asn Trp Val Ala Ala Leu Gly Ser Asp Asp Glu Tyr
        355                 360                 365
Gly Arg Gln Gly Leu Ser Ile Phe Ser Ala Leu Ala Ala Ala Arg Gly
    370                 375                 380
Ile Cys Ile Ala His Glu Gly Leu Val Pro Leu Pro Arg Ala Asp Asp
385                 390                 395                 400
Ser Arg Leu Gly Lys Val Gln Asp Val Leu His Gln Val Asn Gln Trp
                405                 410                 415
Gly Ser Tyr Gly Gly Leu Ala Arg Gly Ile Cys Ala Val Pro Arg Asp
            420                 425                 430
Arg Gln Gln Ser Val Trp Ile Ser Leu Ile Gly Ser Tyr Gly Asp Tyr
        435                 440                 445
Gly Gln Leu Gly Val Gln Ala Leu Glu Glu Leu Ala Val Pro Arg Gly
    450                 455                 460
Ile Cys Val Ala Phe Lys Asp Ile Val Pro Phe Ser Ala Arg Val Gly
465                 470                 475                 480
Asp Pro Arg Met Gln Ser Met Met Gln His Leu Ala Gln Gln Ser Ser
                485                 490                 495
Val Gln Val Val Leu Leu Phe Ala Ser Val His Ala Ala His Ala Leu
            500                 505                 510
Phe Asn Tyr Ser Ile Ser Ser Arg Leu Ser Pro Lys Val Trp Val Ala
        515                 520                 525
Ser Glu Ala Trp Leu Thr Ser Asp Leu Val Met Gly Leu Pro Gly Met
    530                 535                 540
Ala Gln Met Gly Thr Val Leu Gly Phe Leu Val Val Phe His Ala Phe
545                 550                 555                 560
Leu Lys Val Trp Val Ala Ser Glu Trp Ser Gly Gly Thr Val Leu Gly
                565                 570                 575
Ala Arg Thr Thr Val Val Val Val Phe Ser Asn Arg His Leu Ala Arg
            580                 585                 590
Val Phe Phe Arg Ser Val Val Leu Ala Asn Leu Thr Gly Lys Val Trp
        595                 600                 605
Val Ala Ser Glu Asp Trp Ala Ile Ser Thr Tyr Ile Thr Ser Val Thr
    610                 615                 620
Gly Ile Gln Gly Ile Gly Thr Val Leu Gly Val Ala Val Gln Gln Arg
625                 630                 635                 640
Gln Val Pro Gly Leu Lys Glu Phe Glu Glu Ser Tyr Val Arg Ala Val
                645                 650                 655
```

```
Thr Ala Ala Pro Ser Ala Cys Pro Glu Gly Ser Trp Cys Ser Thr Asn
            660                 665                 670
Gln Leu Cys Arg Glu Cys His Thr Phe Thr Thr Arg Asn Met Pro Thr
        675                 680                 685
Leu Gly Ala Phe Ser Met Ser Ala Ala Tyr Arg Val Tyr Glu Ala Val
    690                 695                 700
Tyr Ala Val Ala His Gly Leu His Gln Leu Leu Gly Cys Thr Ser Glu
705                 710                 715                 720
Ile Cys Ser Arg Gly Pro Val Tyr Pro Trp Gln Leu Leu Gln Gln Ile
                725                 730                 735
Tyr Lys Val Asn Phe Leu Leu His Glu Asn Thr Val Ala Phe Asp Asp
            740                 745                 750
Asn Gly Asp Thr Leu Gly Tyr Tyr Asp Ile Ile Ala Trp Asp Trp Asn
        755                 760                 765
Gly Pro Glu Trp Thr Phe Glu Ile Ile Gly Ser Ala Ser Leu Ser Pro
    770                 775                 780
Val His Leu Asp Ile Asn Lys Thr Lys Ile Gln Trp His Gly Lys Asn
785                 790                 795                 800
Asn Gln Val Pro Val Ser Val Cys Thr Thr Asp Cys Leu Ala Gly His
                805                 810                 815
His Arg Val Val Val Gly Ser His His Cys Cys Phe Glu Cys Val Pro
            820                 825                 830
Cys Glu Ala Gly Thr Phe Leu Asn Met Ser Glu Leu His Ile Cys Gln
        835                 840                 845
Pro Cys Gly Thr Glu Glu Trp Ala Pro Lys Glu Ser Thr Thr Cys Phe
    850                 855                 860
Pro Arg Thr Val Glu Phe Leu Ala Trp His Glu Pro Ile Ser Leu Val
865                 870                 875                 880
Leu Ile Ala Ala Asn Thr Leu Leu Leu Leu Leu Leu Val Gly Thr Ala
                885                 890                 895
Gly Leu Phe Ala Trp His Phe His Thr Pro Val Val Arg Ser Ala Gly
            900                 905                 910
Gly Arg Leu Cys Phe Leu Met Leu Gly Ser Leu Val Ala Gly Ser Cys
        915                 920                 925
Ser Phe Tyr Ser Phe Phe Gly Glu Pro Thr Val Pro Ala Cys Leu Leu
    930                 935                 940
Arg Gln Pro Leu Phe Ser Leu Gly Phe Ala Ile Phe Leu Ser Cys Leu
945                 950                 955                 960
Thr Ile Arg Ser Phe Gln Leu Val Ile Ile Phe Lys Phe Ser Thr Lys
                965                 970                 975
Val Pro Thr Phe Tyr Arg Thr Trp Ala Gln Asn His Gly Ala Gly Leu
            980                 985                 990
Phe Val Ile Val Ser Ser Thr Val His Leu Leu Ile Cys Leu Thr Trp
        995                 1000                1005
Leu Val Met Trp Thr Pro Arg Pro Thr Arg Glu Tyr Gln Arg Phe Pro
    1010                1015                1020
His Leu Val Ile Leu Glu Cys Thr Glu Val Asn Ser Val Gly Phe Leu
1025                1030                1035                1040
Leu Ala Phe Thr His Asn Ile Leu Leu Ser Ile Ser Thr Phe Val Cys
                1045                1050                1055
Ser Tyr Leu Gly Lys Glu Leu Pro Glu Asn Tyr Asn Glu Ala Lys Cys
            1060                1065                1070
Val Thr Phe Ser Leu Leu Leu Asn Phe Val Ser Trp Ile Ala Phe Phe
        1075                1080                1085
Thr Met Ala Ser Ile Tyr Gln Gly Ser Tyr Leu Pro Ala Val Asn Val
    1090                1095                1100
Leu Ala Gly Leu Thr Thr Leu Ser Gly Gly Phe Ser Gly Tyr Phe Leu
1105                1110                1115                1120
Pro Lys Cys Tyr Val Ile Leu Cys Arg Pro Glu Leu Asn Asn Thr Glu
                1125                1130                1135
```

His Phe

<210> 5
<211> 101
<212> DNA
<213> HUMAN

<400> 5
gccatctgtg caatgtgcag gtgccagttg catgtgccat gcgtgttggc ygtgagcgtg 60
tgctgttttc gtgtatgtgc catgcacgta tgtgctgcgt g 101

<210> 6
<211> 101
<212> DNA
<213> HUMAN

<400> 6
attcgtgtgc accgtacacg tgtgccattt gcatgtatgc tgtgcacgtg yggcatgcat 60
gtgtgccgtt tgcatgccat gcatgtgttc cttgcgtgtg c 101

<210> 7
<211> 101
<212> DNA
<213> HUMAN

<400> 7
acgtgcggca tgcatgtgtg ccgtttgcat gccatgcatg tgttccttgc rtgtgccgtg 60
cgtgtcccat gcacgtgtgc cgtgcatgtg ccattcgcgt g 101

<210> 8
<211> 101
<212> DNA
<213> HUMAN

<400> 8
cgcaggagtg gcccagacgg gggtcagcag tgtgagtaca gctggccatg yggttcctac 60
agcttccagg cgtcagactc tggcagaagg gctgagaccc t 101

<210> 9
<211> 101
<212> DNA
<213> HUMAN

<400> 9
ggacagagcc caggtggggt gggggcgggg tccagcacca cggccagcac ygaccaccag 60
gaccccggag ccagcaccat ggacagaaaa ctgcccacca g 101

<210> 10
<211> 101
<212> DNA
<213> HUMAN

<400> 10
cctgtcaatt ccgggggggac ggcagccagc tcccggacag ggggtccccc rggtggcccc 60
cccaccactg tataggcctt ggtcgtgggg tggggcgggg g 101

INTERNATIONAL SEARCH REPORT

Int onal Application No
PCT/US 01/07832

A. CLASSIFICATION OF SUBJECT MATTER
IPC 7 C07K14/705 C07K16/28 C12N15/12 A01K67/02 C12P21/00
C12Q1/68 A61K38/17

According to International Patent Classification (IPC) or to both national classification and IPC

B. FIELDS SEARCHED

Minimum documentation searched (classification system followed by classification symbols)
IPC 7 C07K

Documentation searched other than minimum documentation to the extent that such documents are included in the fields searched

Electronic data base consulted during the international search (name of data base and, where practical, search terms used)
EPO-Internal, SEQUENCE SEARCH

C. DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication, where appropriate, of the relevant passages	Relevant to claim No.
E	WO 01 64882 A (MILLENIUM PHARMACEUTICALS) 7 September 2001 (2001-09-07) the whole document, especially figure 15 (seq.id. no. 13 and 15) ---	1-16, 19-23
X	DATABASE EM_HTG 'Online! embl; 13 February 2000 (2000-02-13) LEONGAMONLERT, D.: "Human DNA from clone RP5-89003" retrieved from EBI, accession no. AL139287 Database accession no. AL139287 XP002182829 99,9% identiy in 1814 nt overlap (1625-3438:69303-71116) abstract --- -/--	4-6,8,9

[X] Further documents are listed in the continuation of box C.

[X] Patent family members are listed in annex.

° Special categories of cited documents :

"A" document defining the general state of the art which is not considered to be of particular relevance

"E" earlier document but published on or after the international filing date

"L" document which may throw doubts on priority claim(s) or which is cited to establish the publication date of another citation or other special reason (as specified)

"O" document referring to an oral disclosure, use, exhibition or other means

"P" document published prior to the international filing date but later than the priority date claimed

"T" later document published after the international filing date or priority date and not in conflict with the application but cited to understand the principle or theory underlying the invention

"X" document of particular relevance; the claimed invention cannot be considered novel or cannot be considered to involve an inventive step when the document is taken alone

"Y" document of particular relevance; the claimed invention cannot be considered to involve an inventive step when the document is combined with one or more other such documents, such combination being obvious to a person skilled in the art.

"&" document member of the same patent family

Date of the actual completion of the international search	Date of mailing of the international search report
15 November 2001	28/11/2001

Name and mailing address of the ISA	Authorized officer
European Patent Office, P.B. 5818 Patentlaan 2 NL – 2280 HV Rijswijk Tel. (+31–70) 340–2040, Tx. 31 651 epo nl, Fax: (+31–70) 340–3016	De Kok, A

Form PCT/ISA/210 (second sheet) (July 1992)

C.(Continuation) DOCUMENTS CONSIDERED TO BE RELEVANT

Category °	Citation of document, with indication,where appropriate, of the relevant passages	Relevant to claim No.
X	DATABASE EM_EST 'Online! EMBL; 12 July 1999 (1999-07-12) STRAUSBERG, R.: "Human cDNA clone IMAGE:2419799 3' end" retrieved from EBI, accession no. AI818485 Database accession no. AI818485 XP002182830 94% identity in 411 nt overlap (3458-3048:15-425) abstract	4-6,8,9
A	WO 00 06592 A (UNIVERSITY OF CALIFORNIA) 10 February 2000 (2000-02-10) the whole document, especially pages 79-80 and claim 5	1-16, 19-23
A	US 5 763 569 A (BROWN EDWARD M ET AL) 9 June 1998 (1998-06-09) the whole document, especially figure 47	1-16, 19-23
A	HOON M A ET AL: "PUTATIVE MAMMALIAN TASTE RECEPTORS: A CLASS OF TASTE-SPECIFIC GPCRS WITH DISTINCT TOPOGRAPHIC SELECTIVITY" CELL, MIT PRESS, CAMBRIDGE, MA,, US, vol. 96, 19 February 1999 (1999-02-19), pages 541-551, XP000922524 ISSN: 0092-8674 the whole document	1-16, 19-23

INTERNATIONAL SEARCH REPORT

Information on patent family members

International Application No
PCT/US 01/07832

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
WO 0164882 A	07-09-2001	WO 0164882 A2	07-09-2001
WO 0006592 A	10-02-2000	AU 5238199 A	21-02-2000
		BR 9912545 A	30-10-2001
		CN 1317010 T	10-10-2001
		EP 1100810 A1	23-05-2001
		NO 20010363 A	26-03-2001
		WO 0006592 A1	10-02-2000
US 5763569 A	09-06-1998	US 6011068 A	04-01-2000
		US 5962314 A	05-10-1999
		AU 709303 B2	26-08-1999
		AU 4195796 A	15-05-1996
		BR 9509411 A	30-12-1997
		CA 2202879 A1	02-05-1996
		CN 1220658 A	23-06-1999
		CZ 9701182 A3	17-09-1997
		EP 0787122 A2	06-08-1997
		HU 77980 A2	01-02-1999
		JP 11130737 A	18-05-1999
		JP 2882882 B2	12-04-1999
		JP 10513436 T	22-12-1998
		NZ 297157 A	30-08-1999
		PL 319812 A1	01-09-1997
		WO 9612697 A2	02-05-1996
		US 6001884 A	14-12-1999
		US 5858684 A	12-01-1999
		US 6031003 A	29-02-2000
		US 5688938 A	18-11-1997
		US 6211244 B1	03-04-2001
		AU 3122699 A	22-07-1999
		AU 702629 B2	25-02-1999
		AU 8087294 A	08-05-1995
		CA 2173747 A1	27-04-1995
		CN 1139917 A	08-01-1997
		EP 0724561 A1	07-08-1996
		JP 9504032 T	22-04-1997
		WO 9511221 A1	27-04-1995
		SG 52796 A1	28-09-1998
		WO 9304373 A1	04-03-1993
		AU 673500 B2	14-11-1996
		AU 711247 B2	07-10-1999
		AU 7197796 A	20-02-1997
		CA 2115828 A1	04-03-1993
		CN 1071333 A ,B	28-04-1993
		EP 0657029 A1	14-06-1995
		IL 102917 A	06-12-2000
		JP 2860285 B2	24-02-1999
		JP 9328420 A	22-12-1997
		JP 2887201 B2	26-04-1999
		JP 9281109 A	31-10-1997
		JP 11221095 A	17-08-1999
		JP 2728564 B2	18-03-1998
		JP 6510531 T	24-11-1994
		JP 2001220356 A	14-08-2001
		KR 251995 B1	01-06-2000
		MX 9204881 A1	01-04-1993
		NO 940581 A	25-04-1994

INTERNATIONAL SEARCH REPORT

nformation on patent family members

Int ınal Application No
PCT/US 01/07832

Patent document cited in search report	Publication date	Patent family member(s)	Publication date
US 5763569 A		RU 2147574 C1	20-04-2000